Ph.D. Thesis

Stochastic Comparisons and Bayesian Inference in Software Reliability

Author:
Nuria Torrado Robles

Advisors:
Rosa E. Lillo Rodríguez
Michael P. Wiper

Department of Statistics
Universidad Carlos III de Madrid

Getafe, April 2011

Universidad
Carlos III de Madrid

Ordenaciones Estocásticas e Inferencia Bayesiana en Fiabilidad de Software

Autor:
Nuria Torrado Robles

Memoria presentada para optar al grado de doctor por el programa de **Ingeniería Matemática**, realizada bajo la dirección de:

Dr. Rosa E. Lillo Rodríguez - Catedrática de universidad
Dr. Michael P. Wiper - Titular de universidad

Departamento de Estadística
Universidad Carlos III de Madrid

Getafe, Abril 2011

A mi "pequeña" familia

Agradecimientos

Quiero expresar todo mi agradecimiento a aquellas personas que de alguna forma u otra han facilitado mi andadura hasta el final de este camino, y que probablemente continuarán ayudándome en la siguiente etapa de mi carrera como investigadora.

En primer lugar, quisiera agradecer a los profesores Rosa E. Lillo y Michael P. Wiper por guiarme y enseñarme durante los cuatro años en los que hemos trabajado, codo con codo, para sacar adelante esta tesis. No sólo he aprendido sobre los temas de investigación aquí tratados, sino también sobre su experiencia personal en el mundo académico, lo cuál me servirá en la toma de decisiones futuras. Gracias por darme la oportunidad de asistir a congresos dónde he conocido investigadores de reconocido prestigio, por haber realizado una minuciosa labor de revisión de esta tesis y de los artículos derivados de ella, y sobre todo, por vuestra paciencia y por haber depositado en mí la confianza que a veces a mí misma me falta.

Quisiera agradecer al profesor Simon Wilson que trabajara conmigo durante tres meses durante el otoño del 2008 en Dublín, sus conocimientos y experiencia me ayudaron a comprender mejor algunos métodos computacionales.

También agradezco al profesor Subhash Kochar que aceptara mi visita a Portland State University en el verano del 2010 y que dedicara parte de su tiempo de vacaciones a discutir conmigo algunos temas de investigación presentados en esta tesis.

Quiero mostrar mi gratitud a todos los miembros del departamento de Estadística por permitirme realizar mi tesis doctoral rodeada de investigadores de reconocido prestigio en distintas áreas de Estadística y de Investigación Operativa, y por proporcionarme la financiación necesaria para llevar a cabo esta difícil tarea. En especial, quiero agradecer su cercanía y apoyo a todos los profesores del campus de Leganés. A mis compañeros de doctorado que ya terminaron, Alba, Ángel, Lee, Omar, Pepa, Peter, por compartir sus experiencias conmigo; a los que aún siguen, Ana Paula, Cristina, Dalia, Joanna, Miguel Ángel, Sofía y todos los demás, mucho ánimo. Y en especial a mi compañero de despacho en Leganés, Henry.

Finalmente quiero agradecer a toda mi familia, sanguínea y política, por estar a mi lado ahora y siempre. A mis padres por enseñarme que el esfuerzo es parte de la vida, y que así, se puede llegar hasta dónde uno quiere. A Maruja, por ser la primera persona en llamarme "doctora" mucho tiempo antes de serlo. A mi hermana, por ser mi mejor amiga, mi confidente y mi apoyo en momentos difíciles. Y a Edy, la persona más importante de mi vida, a quién le dedico esta memoria. Gracias por tu cariño y por no dejarme dar un paso atrás ni para tomar carrerilla.

qed: *quod erat demostrandum*

Abstract

Within the last decade of the 20th century and the first few years of the 21st century, the demand for complex software systems has increased, and therefore, the reliability of software systems has become a major concern for our modern society. Software reliability is defined as the probability of failure free software operations for a specified period of time in a specified environment. Many current software reliability techniques and practices are detailed by Lyu [54] and Pham [77].

From a statistical point of view, the random variables that characterize software reliability are the epoch times in which a failure of software takes place or the times between failures. Most of the well known models for software reliability are centered around the interfailure times or the point processes that they generate. A software reliability model specifies the general form of the dependence of the failure process on the principal factors that affect it: fault introduction, fault removal, and the operational environment.

The purpose of this thesis is threefold: (1) to study stochastic properties of times between failures relative to independent but not identically distributed random variables; (2) to investigate properties of the epoch times of nonhomogeneous pure birth processes as an extension of nonhomogeneous Poisson processes used in the literature in software reliability modelling and, (3) to develop a software reliability model based on the use of covariate information such as software metrics. Firstly, properties of statistics based on heterogeneous samples will be investigated with the aid of stochastic orders. Stochastic orders between probability distributions is a widely studied concept. There are several kinds of stochastic orders that are used to compare different aspects of probability distributions like location, variability, skewness, dependence, etc. Secondly, ageing notions and stochastic orderings of the epoch times of nonhomogeneous pure birth processes are studied. Ageing notions are another important concepts in reliability theory. Many classes of life distributions are characterized or defined according to their aging properties in the literature. Finally, we exhibit a non-parametric model based on Gaussian processes to predict number of software failures and times between failures. Gaussian processes are a flexible and attractive method for a wide variety of supervised learning problems, such as regression and classification in machine learning.

This thesis is organized as follows. In Chapter 1, we present some basic software reliability measures. After providing a brief review of stochastic point processes and models of ordered random variables, it discusses the relationship between these kind of models and types of failure data. This is then followed by a brief review of some stochastic orderings and ageing notions. The chapter concludes with a review of some

well known software reliability models.

The results of Chapter 2 concern stochastic orders for spacings of the order statistics of independent exponential random variables with different scale parameters. These results on stochastic orderings and spacings are based on the relation between the spacings and the times between successive software failures. Due to the complicated expression of the distribution in the non-iid case, only limited results are found in the literature. In the first part of this chapter, we investigate the hazard rate ordering of simple spacings and normalized spacings of a sample of heterogeneous exponential random variables. In the second part of this chapter, we study the two sample problem. Specifically, we compare both simple spacings and normalized spacings from two samples of heterogeneous exponential random variables according to the likelihood ratio ordering. We also show applications of these results to multiple-outlier models.

In Chapter 3, motivated by the equality in distribution between sequential order statistics and the first n epoch times of a nonhomogeneous pure birth process, we consider the problem of comparing the components of sequential k-out-of-n systems according to magnitude and location orders. In particular, this chapter discusses conditions on the underlying distribution functions on which the sequential order statistics are based, to obtain ageing notions and stochastic comparisons of sequential order statistics. We also present a nonhomogeneous pure birth process approach to software reliability modelling.

A large number of models have been proposed in the literature to predict software failures, but a few incorporate some significant metrics data observed in software testing. In Chapter 4, we develop a new procedure to predict both interfailure times and numbers of software failures using metrics information, from a Bayesian perspective. In particular, we develop a hierarchical non-parametric regression model based on exponential interfailure times or Poisson failure counts, where the rates are modeled as Gaussian processes with software metrics data as inputs, together with some illustrative concrete examples.

In Chapter 5 we show some general conclusions and describe the most significant contributions of this thesis.

RESUMEN

En la última década del siglo 20 y en los primeros años del siglo 21, la demanda de sistemas informáticos ha aumentado considerablemente, muestra de ello es su presencia en satélites espaciales, aviones, cadenas de montaje automatizadas, incluso cada vez están más cercanos a nuestra vida cotidiana como en automóviles, electrodomésticos o teléfonos móviles. Un sistema informático consta de dos tipos de componentes: el hardware y el software. Entre ellos la principal diferencia es que el software no se desgasta. Así, un programa informático podría funcionar al cabo de años con la misma corrección con que lo hizo el primer día sin necesidad de modificación alguna.

En general, la calidad de un producto puede valorarse desde diversos puntos de vista. El software no es una excepción, y existen por tanto diferentes enfoques para la valoración de su calidad. Aquí nos centraremos en uno de dichos enfoques: la *fiabilidad*. Por fiabilidad se entiende la probabilidad de ausencia de fallos durante la operación de un producto de software. Existen diferentes técnicas estadísticas para medir la fiabilidad de un programa informático, algunas de ellas son detalladas en Lyu [54] y Pham [77].

Desde un punto de vista estadístico, las variables aleatorias que caracterizan la fiabilidad del software son los instantes de tiempo en los que se produce un fallo de software, así como, los tiempos entre fallos. Uno de los objetivos principales de esta tesis es modelizar el comportamiento de dichas variables aleatorias. Resulta interesante estudiar el comportamiento estocástico de dichas variables, ya que, de este modo, podemos conocer propiedades de las mismas relacionadas con sus funciones de supervivencia o con sus funciones de tasa de fallo. En este sentido, en el Capítulo 2, presentamos resultados referidos con ordenaciones estocásticas de los tiempos entre fallos de software, relativos a variables aleatorias independientes no idénticamente distribuidas. Estos resultados se basan en la relación que liga dichos tiempos con los espaciamientos (*spacings*). Tanto los estadísticos de orden como los espaciamientos tienen un gran interés en el contexto del Análisis de Supervivencia, así como en la Teoría de Fiabilidad. En la mayoría de los trabajos existentes, se asume que las variables implicadas son independientes e idénticamente distribuidas (iid). Debido a la complejidad analítica que conlleva relajar alguna de estas dos hipótesis, no hay demasiadas referencias para el caso en el que las variables no sean iid. Kochar y Korwar [43] comprobaron que, cuando el número de exponenciales que se contemplan son tres, los espaciamientos normalizados cumplen la ordenación de tasa de fallo y conjeturaron lo mismo para el caso general de n variables aleatorias exponenciales heterogéneas. En la Sección 2.2, se presentan avances relacionados con dicha conje-

tura, así como, resultados relativos a la ordenación de tasa de fallo de espaciamientos sin normalizar. También han sido estudiados en este capítulo problemas asociados con espaciamientos obtenidos a partir de muestras aleatorias de dos poblaciones. En particular, hemos obtenido condiciones suficientes para que se verifique la ordenación de razón de verosimilitud entre espaciamientos de dos muestras de exponenciales heterogéneas.

Por otra parte, hemos trabajado con estadísticos de orden secuenciales, ya que incluyen un gran número de variables aleatorias ordenadas. Además, este tipo de estadísticos de orden son interesantes porque están ligados con los tiempos en los que ocurre un fallo de procesos no homogéneos de nacimiento puro. Cabe destacar, que este tipo de variables son dependientes y no idénticamente distribuidas, lo que aumenta la complejidad del problema. Nuestro objetivo aquí, es estudiar qué condiciones deben verificar las distribuciones subyacentes a partir de las cuales se definen los estadísticos de orden secuenciales para que éstos cumplan algún tipo de ordenación estocástica. Los resultado obtenidos en este sentido se presentan en el Capítulo 3. En este capítulo, también estudiamos otro concepto importante en fiabilidad, la noción de envejecimiento. Los diferentes conceptos de envejecimiento describen como una componente o un sistema mejora o empeora con la edad. En este sentido, el envejecimiento positivo significa que las componentes tienden a empeorar debido al desgaste. Exactamente esto es lo que le ocurre al hardware. Mientras que, cuando un sistema supera ciertos tests y mejora, diremos que el envejecimiento es negativo, como le sucede al software. En el segundo capítulo de la tesis, estudiamos condiciones bajo las cuales algunas propiedades de envejecimiento verificadas por las distribuciones subyacentes, a partir de las cuales se definen los estadísticos de orden secuenciales, se cumplen también para los estadísticos de orden secuenciales.

Si bien es cierto que se han desarrollado en los últimos cuarenta años un gran número de modelos de fiabilidad de software, la mayoría de ellos no tienen en consideración la información proporcionada por covariables. Otra aportación de esta tesis, la cual se encuentra en el Capítulo 4, consiste en la utilización de métricas del software como variables independientes para predecir o bien el número de fallos de un programa informático o bien los tiempos entre sucesivos fallos del software. Una métrica de un programa informático sirve para medir la complejidad y la calidad del mismo, así como, la productividad de los programadores con respecto a su eficiencia y competencia. En esta tesis, hacemos uso de métricas para medir la complejidad de un programa informático a través del volumen del mismo contabilizando el número de líneas de código. En la literatura existen algunos modelos lineales para predecir datos de fallos del software mediante métodos de inferencia clásicos. Sin embargo, nosotros

optamos por utilizar procesos gaussianos que relajan la linealidad y que han sido ampliamente usados en problemas de aprendizaje automático, tanto en regresión como en clasificación.

Por último, en el Capítulo 5, resumimos las principales aportaciones de esta tesis.

Contents

List of Tables

List of Figures

List of Acronyms and Notations

A

AIC Akaike information criterion, p. 93.

B

BIC Bayesian information criterion, p. 93.

C

CLOC Comments and blank lines of code, p. 80.

D

DLR Decreasing likelihood ratio, p. 15.

DHR Decreasing hazard rate, p. 15.

DHRA Decreasing hazard rate average, p. 15.

DIC Deviance information criterion, p. 93.

DRHR Decreasing reversed hazard rate, p. 15.

DU Duane model, p. 19.

E

EOS Exponential order statistics, p. 9.

G

GO Goel-Okumoto model, p. 19.

GOOS General Order statistics, p. 9.

GP Gaussian process, p. 87.

H

$h(t)$ Hazard (or failure) rate function, p. 3.

$H(t)$ Cumulative hazard rate function, p. 4.

HPBP Homogeneous pure birth process, p. 20.

I

IHR Increasing hazard rate, p. 15.

IRHR Increasing reversed hazard rate, p. 15.

IHRA Increasing hazard rate average, p. 15.

ILR Increasing likelihood ratio, p. 15.

$\lambda(t)$ Intensity function of a counting process, p. 6.

J

JM Jelinski-Moranda model, p. 17.

K

K&K Kochar and Kowar, p. 27.

L

LOC Lines of code, p. 80.

LV Littlewood-Verrall model, p. 85.

M

MCMC	Markov Chain Monte Carlo, p. 82.
$\Lambda(t)$	Mean value function of a counting process, p. 6.
MO	Musa-Okumoto model, p. 20.

N

NCLOC	Non-comment lines of code, p. 80.
NCNCSL	New or changed non-commentary source lines, p. 94.
NHPB	Nonhomogeneous pure birth process, p. 7.
NHPP	Nonhomogeneous Poisson process, p. 6.

O

OOS	Order statistics, p. 8.
$X_{i:n}$	i'th order statistic, p. 8.

P

PIM$(\cdot,\cdot)$	Proportional intensity model, p. 94.
PSE	Prediction square error, p. 93.

R

$r(t)$	Reversed hazard rate function, p. 5.

S

$X^*_{i:n}$	i'th sequential order statistic, p. 11.
SOS	Sequential order statistics, p. 10.
$D_{i:n}$	i'th simple spacing, p. 9.
$D^*_{i:n}$	i'th normalized spacing, p. 26.
SRM	Software reliability model, p. 15.
$\leq_{st}$	usual stochastic order, p. 13.

$\leq_{hr}$ Hazard rate order, p. 13.

$\leq_{rh}$ Reversed hazard rate order, p. 13.

$\leq_{lr}$ Likelihood ratio order, p. 13.

$\leq_{\text{disp}}$ Dispersive order, p. 14.

$\leq_{*}$ Star order, p. 14.

$\leq^{m}$ Majorization order, p. 14.

$\overline{F}(t)$ Survival function, p. 3.

SW Shick-Wolverton model, p. 18.

T

S_i i'th time to software failure, p. 6.

T_i i'th time between successive software failures, p. 6.

TP2 Totally positive of order 2, p. 67.

CHAPTER 1

Introduction

The central theme of this thesis is the study of reliable software systems, from a statistical point of view, as a key element of computer systems. Today, computer systems permeate our modern society. They are embedded in air traffic control, nuclear reactors, aircraft, real-time sensor networks, industrial process control, automotive mechanical and safety control, and hospital health care, among others. As the functionality of computing operations becomes more essential, there is a greater need for a high reliability of the computing systems. They combine both software and hardware that have to function together to complete various tasks.

In order to explain the interest of the topics of this dissertation, we present a real example, the Year 2000 Problem (Y2K). An impressive list of system crashes due to software and their cost is reported in Lyu [54] and Pham [77]. In the early 1970s, when computers were first used in the business world, storage space was at a premium and the use of a two-digit convention to represent the year seemed appropriate. Incorrect software programs will assume that the maximum value of a year field is "99"and will roll systems over to the year 1900 instead of 2000, resulting in negative date calculations and the creation of many overnight centenarians. Some consequences of the Y2k problem verged on the serious, like the glitches that hit the Japanese nuclear power plants and the US military satellite, but most were mundane, from broken bus ticket machines in Tasmania to police breath-testing equipment in Hong Kong (BBC, January 4, 2001).

The year 2000 was, however, not the only date dangerous to software applications. Some systems had problems that rolled over to 2010. The most important such

problems occurred in Germany, where around a quarter of debit and credit cards have been rendered unreadable by the software bug, causing chaos across the country. More than 20 million of these banks' cards were not working when the glitch cropped up on New Year's Day (Time, January 7, 2010). Another case was that of PlayStation 3 consoles. A calendar-based error afflicting older PlayStation 3 consoles has meant that owners worldwide are hesitant to turn on their gaming machines until Sony assures them that a fix is available. This error was due to reset its calendar to December 31, 1999 in America and the January 1, 2000 in Europe and Asia. (The Independent, March 1, 2010).

These examples highlight that the development of reliable software programs is a necessity today. Software reliability is an important metric to assess the correct functioning of software systems. This chapter defines software reliability models and illustrates the prime importance of such techniques in quantifying the reliability of repairable systems. We give a brief description of different reliability measures and review some very well known stochastic orders, based on the comparison of these reliability measures. Also, we provide a short introduction to counting processes and models of ordered random variables, because these two types of models have been applied to software reliability models in the literature.

1.1. SOFTWARE RELIABILITY MEASURES

Reliability is defined as the probability that a system will perform its intended function under specified design limits. A computer system consists of two major components: hardware and software. In general, a system can be defined as a collection of two or more parts which is designed to perform one or more functions. A software system is a *repairable system*, i.e., it can be restored to fully satisfactory performance by any method, other than replacement of the entire system, after failing to perform one or more of its functions satisfactorily. It is obvious that most real world systems are repairable systems.

Software reliability is different from hardware reliability in the sense that software does not wear out or burn out. The software itself does not fail unless flaws within the software result in a failure in its dependent system. Furthermore, software systems are usually debugged during testing phase so that its reliability is improving over time. Software reliability is an attribute and key factor in software quality (Lyu [54]) and is defined as follows.

Definition 1.1.1. *Software reliability* is the probability of failure-free software operation for a specified period of time in a specified environment.

Three major components in the definition of software reliability are *failure, time* and *operational environment*. The *operational environment* of a system is a group of runs which typically involve similar processing.

A *failure* occurs when the user perceives that the program ceases to deliver the expected service. A *fault* is uncovered when either a failure of the program occurs or an internal error is detected within the program. It is noteworthy that all faults do not necessarily cause failures, but all failures are caused by faults. The causes of software failure are different from those of hardware failure. A consequence is that it is possible to have software that is fault-free and so will never experience failure for any mission time, whereas hardware experiences deterioration with use. Software fails because of faults in the code and these faults are introduced due to human error.

Reliability quantities are defined with respect to time. We are concerned with three types of time: (1) the *execution time* for a software system is the CPU time that is actually spent by the computer in executing the software; (2) the *calendar time* is the time people normally experience in terms of years, months, etc; and (3) the *clock time* is the elapsed time from start to end of computer execution in running the software.

Once time basis is determined, failures can be expressed in several ways: the survival function, the hazard rate function and the reversed hazard rate function. We introduce below the definitions of these three well known reliability measures. Let X be the random variable representing the time to a failure or the lifetime of a system and let F denote the cumulative distribution function of X. Then F represents the probability that a failure occurs at or before time t. The survival function of X can be defined as follows.

Definition 1.1.2. Let X be a random variable, if F denotes the distribution function of X, then $\overline{F}(t) = 1 - F(t)$ denotes the corresponding *survival function*. That is,

$$\overline{F}(t) = P(X > t).$$

The hazard rate function is a measure of the tendency to fail, it is also known as the instantaneous failure rate.

Definition 1.1.3. The *hazard* or *failure rate function*, $h(t)$, of a random variable X at t is defined on the support of the distribution by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < X \leq t + \Delta t \mid X > t)}{\Delta t}. \tag{1.1.1}$$

If the density function, $f(t)$, of X exists, then the hazard rate function can alternatively be expressed as

$$h(t) = \frac{f(t)}{\overline{F}(t)}. \tag{1.1.2}$$

As can be seen from (1.1.1), the hazard rate function measures the proneness to failure at time t in that

$$h(t)\Delta t \approx P(X \leq t + \Delta t \mid X > t),$$

for small Δt. The *cumulative hazard rate function* is denoted by $H(t)$,

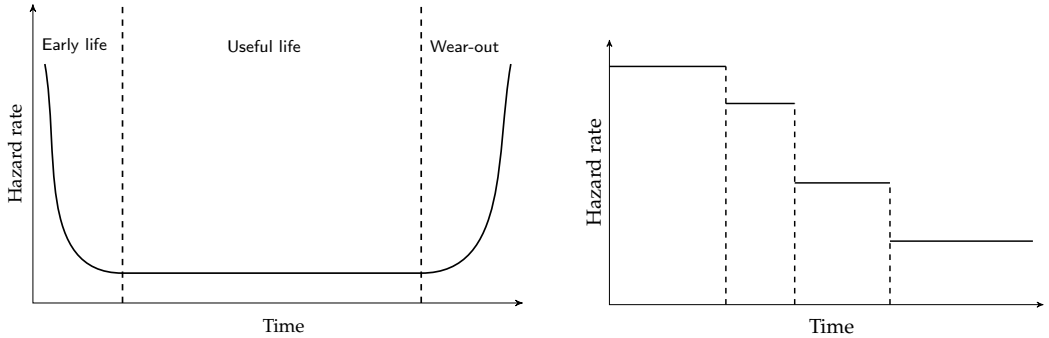
$$H(t) = \int_0^t h(z)dz = -\ln\big(\overline{F}(t)\big). \tag{1.1.3}$$

The well known relation

$$\overline{F}(t) = \exp\big(-H(t)\big), \tag{1.1.4}$$

establishes the link between the cumulative hazard rate function and the survival function.

As we mentioned before, software reliability is different from hardware reliability. This difference can be seen from their respective hazard rate function. There is a well known *bathtub* curve in reliability studies for hardware products. The curve is given in Figure 1.1a. The shape of the curve is like a bath tub. There are three phases for the life of a hardware system. The initial phase is the burn-in phase, where the failure rate is high. It is expected that the product is tested in industry before delivery. Due to testing and fixing faults, the failure rate will come down initially and may stabilize after a certain time. The second phase is the useful life phase where the failure rate is approximately constant and is called useful life of a hardware product. After a few years, again the failure rate will increase due to wearing out of components. This phase is called wear out phase. We do not have this phase for the software as, clearly, it does not wear out. The curve for software is given in Figure 1.1b, where one can be seen that the hazard rate of a software system is generally decreasing due to the discovery and removal of software failures.



(a) Hardware system (b) Software system

Figure 1.1: Hazard rate function for the two major components of a computer system

Definition 1.1.4. If X is a random variable with a distribution function F then the *reversed hazard rate function* of X, $r(t)$, at the point t is defined as

$$r(t) = \lim_{\Delta t \to 0} \frac{P(t - \Delta t \leq X < t \mid X < t)}{\Delta t}. \tag{1.1.5}$$

If F is absolutely continuous distribution function with density function f, then the reversed hazard rate function is defined in a manner similar to the hazard rate function in (1.1.2), but with the distribution function replacing the survival function, that is,

$$r(t) = \frac{f(t)}{F(t)}. \tag{1.1.6}$$

From (1.1.5), $r(t)\Delta t$ can be interpreted as an approximate probability of a failure in $(t - \Delta t, t]$ given that the failure had occurred in $[0, t]$. It is easy to prove that the following analog of the exponential formula (1.1.4) takes place,

$$F(t) = \exp\Big(\int_t^\infty r(z)dz\Big). \tag{1.1.7}$$

Making simple transformations we arrive at an important relation between $h(t)$ and $r(t)$,

$$r(t) = h(t)\frac{\overline{F}(t)}{F(t)} = h(t)\Big(\exp(H(t)) - 1\Big)^{-1}. \tag{1.1.8}$$

For more details see Finkelstein [23].

Two types of failure data, namely, *failure count* data and *time between failures* data, can be collected for the purpose of software reliability measurement. Failure count data tracks the number of failures detected per unit of time and time between failures data tracks the intervals between consecutive failures. We will see in Section 1.2 and Section 1.3 the relationship between these types of failure data and stochastic point processes, and models of ordered random variable, respectively, used in the literature to model these events.

1.2. STOCHASTIC COUNTING PROCESSES

As we will see more in details in Section 1.5, the two main types of models, that have been applied to software systems, are stochastic counting processes and models of ordered random variables, . In this section, we will treat only counting process models, deferring the discussion of models of ordered random variables to the next section.

Counting process models have played a key role in the analysis of software failure data. A *counting process*, $\{N(t), t \geq 0\}$, is simply a count of the number of events that

have occurred in any specified interval of time. In a typical case, the random variable S_i represents the times at which events of a specified character have occurred, and it is called *epoch time* or *waiting time*, for $i = 1, 2, \ldots$. The random variables $T_1 = S_1, T_2 = S_2 - S_1, \ldots, T_i = S_i - S_{i-1}$ are called the successive *interarrival times*. In the context of our mark, S_i is the i'th software failure and T_i is the times between the i'th software failure and the $(i-1)$'th software failure. A typical history of a software program is given in Figure 1.2. We assume that at time zero the program is run on the computer and works satisfactorily until time s_1, when the first failure occurs. The programmer then repairs the program, it works satisfactorily for time t_2, it is repaired again, and so on.

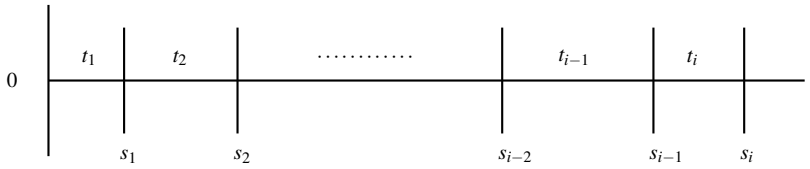


Figure 1.2: Typical failure history of a software program

A counting process is said to possess *independent increments* if the number of events that occur in disjoint time intervals are independent. A counting process is said to possess *stationary increments* if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval.

The *Poisson process* model is one of the simplest and perhaps the best well known of all counting process models. These models have stationary and independent increments. This means that $\{N(t), t \geq 0\}$ is Poisson distributed with mean $\Lambda(t)$ where $\Lambda(t)$ is the *mean value function*. Besides, $\Lambda(t) = \lambda t$, by stationary increments, so that $\Lambda(t)$ is directly proportional to t, with proportionality factor λ, the mean rate at which counts are being made. In other words, the process has no memory, and hence exponential interarrival times are to be expected.

The mean value function of a counting process is always assumed to be continuous and is also assumed to be differentiable, with derivative denoted by $\lambda(t)$, that is,

$$\lambda(t) = \frac{d}{dt}\Lambda(t),$$

which is called the *intensity function*. See Parzen [73] for a good review of stochastic processes.

A Poisson process with non-stationary increments is called a nonhomogeneous Poisson process (NHPP), and this process can be defined as follows.

Definition 1.2.1. A counting process $\{N(t), t \geq 0\}$ is a *nonhomogeneous Poisson process* with intensity function $\lambda(t)$ if

a) $\{N(t), t \geq 0\}$ has the Markov property,

b) $P(N(t+\Delta t) = n+1 \mid N(t) = n) = \lambda(t)\Delta t + o(\Delta t)$, $n \geq 1$,

c) $P(N(t+\Delta t) > n+1 \mid N(t) = n) = o(\Delta t)$, $n \geq 1$.

Let $S_1,\ldots,S_n$ be the successive epoch times of an NHPP, where $S_0 \equiv 0$, then the density function of S_i is,

$$f_i(t) = \lambda(t)\frac{\Lambda^{i-1}(t)}{(i-1)!}e^{-\Lambda(t)}, \quad t \geq 0, \tag{1.2.9}$$

(see (2.3) in Belzunce et al. [8] and (3) in Baxter [7]). We obtain its hazard rate function as

$$h_i(t) = \lambda(t)\frac{\Lambda^{i-1}(t)}{(i-1)!\sum_{j=0}^{i-1}\frac{\Lambda^j(t)}{j!}}, \quad t \geq 0. \tag{1.2.10}$$

The nonhomogeneous Poisson process can be generalized to what can be called a nonhomogeneous pure birth process (NHPB). As we mentioned the intensity of a jump of a nonhomogeneous Poisson process at any time t depends only on t, and not on any other information about the past or the present of the process. However, in a nonhomogeneous birth process the intensity of a jump at any time t depends on both t and the state of the process, that is, the number of previous jumps. The following definition summarizes these conditions.

Definition 1.2.2. A counting process $\{N(t), t \geq 0\}$ is a *nonhomogeneous pure, birth process* with intensity function $\lambda_n(t)$ if

a) $\{N(t), t \geq 0\}$ has the Markov property,

b) $P(N(t+\Delta t) = n+1 \mid N(t) = n) = \lambda_n(t)\Delta t + o(\Delta t)$, $n \geq 1$,

c) $P(N(t+\Delta t) > n+1 \mid N(t) = n) = o(\Delta t)$, $n \geq 1$.

It is known in the literature the relation between the distribution of the jump times of a NHPB process and the distributions of ordered random variables. The next section is devoted to explain such link.

1.3. MODELS OF ORDERED RANDOM VARIABLES

Models of ordered random variables are widely used in statistical modelling and inference. In this section we review some models of ordered random variables: ordinary order statistics, record values and sequential order statistics (cf. Kamps [34]

and [35]). We also discus the relations between these models and stochastic counting process, for more details see Lenz [52].

If the random variables $X_1, \ldots, X_n$ are arranged in ascending order of magnitude, then the i'th smallest of X_i's is denoted by $X_{i:n}$. The ordered quantities

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}\,, \tag{1.3.11}$$

are called *ordinary order statistics* (OOS), and $X_{i:n}$ is the i'th order statistic. These random variables are of great interest in many areas of statistics, in particular, in the characterizations of distributions (cf. Deheuvels [20]) and testing problems (see, e.g., Berrendero et al. [9]), among others. Specifically, there is a very interesting application of OOS's in reliability theory. The $(n-k+1)$'th OOS in a sample of size n represents the life length of a k-out-of-n system which is an important technical structure. It consists of n components of the same kind with independent and identically distributed life lengths. All n components start working simultaneously, and the system works, if at least k components function; i.e. the system fails, if $(n-k+1)$ or more components fail. As an example, we will look at a 2-out-of-3 system. This system can be illustrated by the reliability block diagram shown in Figure 1.3. Special cases of k-out-of-n systems are series and parallel systems as we show in the following examples.

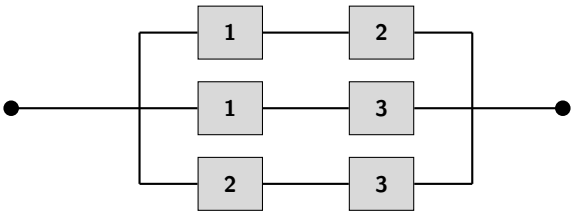


Figure 1.3: A 2-out-of-3 system

Example 1.3.1 A system that is functioning if and only if each component is functioning is called a *series system* and is represented by a n-out-of-n system. Its lifetime is described by the smallest lifetime, $X_{1:n}$. A series structure can be illustrated by the reliability block diagram in Figure 1.4a. The survival function of this system is given by

$$\overline{F}_{1:n}(t) = \prod_{i=1}^{n} \overline{F}_i(t),$$

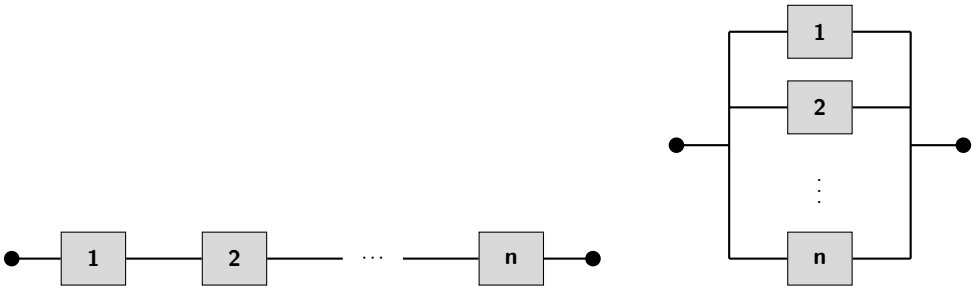
where the X_i's are assumed to be independent and $\overline{F}_i$ is the survival function of X_i, for $i = 1, \ldots, n$. ◁

Example 1.3.2 A system that is functioning if and only if at least one component is functioning is called a *parallel system* and is represented by a 1-out-of-n system. Its

lifetime is described by the largest lifetime, $X_{n:n}$. The corresponding reliability block diagram is shown in Figure 1.4b. The cumulative distribution function of this system is given by

$$F_{n:n}(t) = \prod_{i=1}^{n} F_i(t),$$

where the X_i's are assumed to be independent and F_i is the distribution function of X_i, for $i = 1, \ldots, n$. ◁



(a) Series system (b) Parallel System

Figure 1.4: Special cases of k-out-of-n systems

Another interesting random variables are $D_{i:n} = X_{i:n} - X_{i-1:n}$, when $X_{0:n} \equiv 0$, called *simple spacings*. In the software reliability context they correspond to times elapsed between successive software failures.

Since the times to software failure $0 \equiv S_0 \leq S_1 \leq \cdots \leq S_i \leq \cdots$ are ordered, they constitute a natural framework for an order statistics type analysis. Note that OOS formed from independent random variables are dependent, the act of ordering destroys independence. When $X_1, \ldots, X_n$ are independent but not necessarily identically distributed (inid) exponential random variables, the resulting model is called the *exponential order statistics model* (EOS) in Miller [63]. A special case of the EOS model, when the exponential random variables are identically distributed (iid), is the model by Jelinski and Moranda [33] which we will define in Section 1.5. Different forms for the common cumulative distribution function, say the Pareto, the Weibull, and so on, will lead to different probability models for the epoch of failures. These models have been referred to by Raftery [79] as the *general order statistics models* (GOOS), not to be confused with generalized order statistics models. In the GOOS model we assume that there is an unknown number of faults N at the beginning of software testing. We model the observed failure epochs to be the first n OOS taken from N iid observations. Kuo and Yang [49] studied the relationship between GOOS models and NHPP processes in SR context.

Theorem 1.3.3 (Kuo and Yang [49]). *Suppose that failure epochs are described by a GOOS model with a distribution function $F(t)$ and a parameter N. Let $\{N(t), t \geq 0\}$ denote the number of epochs in time $[0,t]$. Assuming that N has a Poisson distribution with parameter θ, then $\{N(t), t \geq 0\}$ can be described by a nonhomogeneous Poisson process with mean value function $\Lambda(t) = \theta F(t)$.*

Observe that, in this case, we have that $\Lambda(t) < \infty$ as $t \to \infty$. Therefore, GOOS models is limited to testing where no new faults are introduced at each repair. However, record values models can incorporate the situation where new faults may be added during repairs, because $\Lambda(t) \to \infty$ as $t \to \infty$.

Record values are also used in software reliability (see Kuo and Yang [49]), they are defined as a model for successive extremes in a sequence of independent and identically distributed random variables. A good introduction can be found in Arnold et al. [1]. Record values are closely connected with the epoch times of some corresponding nonhomogeneous Poisson process often used in shock models. Gupta and Kirmani [28] showed that the sequence of record values can be viewed as the sequence of epoch times of some nonhomogeneous Poisson process and viceversa, if the distribution function F upon which the records are based, and the mean value function $\Lambda(t)$ of the process satisfy $F(t) = 1 - \exp\Big(-\Lambda(t)\Big)$. They also gave a survey on the application of nonhomogeneous Poisson processes in the modelling of repairable systems. Enlarged and generalized models of record values are discussed in Kamps [34].

There are numerous analogies in the properties and the behaviour of OS's and record values. For example, if the exponential distribution is the underlying distribution, then successive differences of OS's and of record values are independent and again exponentially distributed (Sukhatme [95]).

Kamps [34] introduced the concept of *sequential order statistics* (SOS) as an extension of OOS model. Sequential order statistics model the reliability of certain k-out-of-n systems without the assumption of independence of the lifetime of the components. In this model, the lifetime distribution of the remaining components in the system may change after each failure of the components. At the beginning, the lifetimes of the components are iid with a common distribution function F_1. After the first component fails, the distribution of the residual lifetimes of the remaining $(n-1)$ components changes to that of the residual lifetime distribution of a second distribution F_2. If we observe the i'th failure at time t, the remaining $(n-i)$ components are now supposed to have a possibly different distribution. Proceeding in this way we obtain a triangular scheme of random variables where the i'th line containing $n-i+1$ random variables with distribution function F_i, $1 \leq i \leq n$, indicating that $i-1$ components previously failed. Following Cramer and Kamps [18], sequential order statistics are defined as

follows.

Definition 1.3.4. Let $F_1,\dots,F_n$ be distribution functions with $F_1^{-1}(1) \le \cdots F_n^{-1}(1)$ and let $B_1,\dots,B_n$ independent random variables with $B_i \sim \text{Beta}(n-i+1,1)$, $1 \le i \le n$. Then the random variables

$$X^*_{i:n} = F_i^{-1}\Big(1 - B_i \overline{F}_i(X^*_{i-1:n})\Big), \quad \text{for } i = 1,\dots,n,$$

are called *sequential order statistics*.

Note that OOS are contained in the model of SOS by the specific choice $F_1 = \cdots = F_n$. In the same way that there exists a connection between OOS and the lifetimes of k-out-of-n systems, there exists a relation between SOS and the lifetimes of sequential k-out-of-n systems. In this case, the $(n-k+1)$'th SOS in a sample of size n represents the life length of a sequential k-out-of-n system. A sequential k-out-of-n system is more flexible than a k-out-of-n system in the sense that, after the failure of some component, the distribution of the residual lifetime of the components at work may change.

The model of SOS is closely connected to several other models of ordered random variables. In its general form the model coincides with Pfeifer's record model, see e.g. Pfeifer [76] and Kamps [34]. The specific choice of distribution functions $F_i(t) = 1 - (1-F(t))^{\alpha_i}$, $t \in \Re$, $1 \le i \le n$, with a distribution function F and positive real numbers $\alpha_1,\dots,\alpha_n$ leads to the model of generalized order statistics with parameters $\gamma_i = (n-i+1)\alpha_i$, $1 \le i \le n$.

It is noteworthy that the distribution of sequential order statistics coincides with the distribution of the first n epoch times of a nonhomogeneous pure birth process (NHPB) under some condition, as we show in the following proposition.

Proposition 1.3.5 (Proposition 2.1. in Zhuang and Hu [105])**.** *Let $h_1,\dots,h_n$ be the hazard rate functions of distribution $F_1,\dots,F_n$, respectively, and let $X^*_{1,n},\dots,X^*_{n,n}$ be the SOS based on $\{F_1,\dots,F_n\}$. Define*

$$\lambda_i(t) = (n-i+1)h_i(t), \quad \textit{for } i = 1,\dots,n,$$

and denote by S_i the i'th epoch time of a NHPB with intensity function $\lambda_i(t)$ for $i = 1,\dots,n$. Then

$$\left(X^*_{1,n},\dots,X^*_{n,n}\right) \stackrel{st}{=} (S_1,\dots,S_n),$$

where $\stackrel{st}{=}$ means equality in distribution.

We want to use this equivalence to incorporate NHPB processes in software reliability. In fact, we have not seen any papers on software reliability that use neither SOS models nor NHPB processes. NHPB processes can be appropriate for the following reason. NHPP processes are extensively used to model failures for software systems,

as we will see in Section 1.5, and NHPB processes are a generalization of these processes. Then, we can consider the NHPB process as a unified framework that incorporates all of the software reliability modelling based on NHPP processes. An overview is sketched in Figure 1.5.

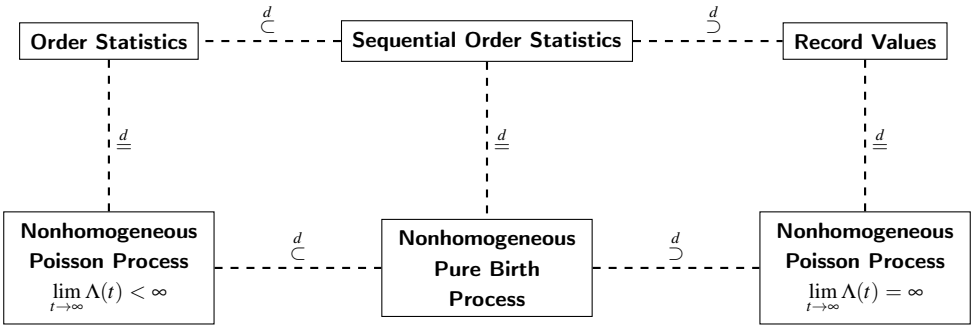


Figure 1.5: Overview of some models of ordered random variables

1.4. STOCHASTIC ORDERINGS AND AGEING NOTIONS

Suppose we want to compare two different distributions. What is the best way to do that? In general, the simplest way to compare two distribution functions is by their associated means and variances or standard deviations. However, such a comparison is not very informative. In addition to this, the means and the variances sometimes do not exist, as for example, for Cauchy distributions. Comparisons based on survival function, hazard rate function and reversed hazard rate function of distributions often establish partial orders among them, which are well known as *stochastic orders*. Since 1994 the theory of stochastic orders has grown significantly, see e.g. Shaked and Shanthikumar [88]. In this thesis, we use the tool of stochastic orders to investigate stochastic comparisons between successive spacings based on OOS from a sample of heterogeneous exponential random variables, and between successive SOS.

Another important concept in reliability theory is the concept of ageing. By ageing we mean "the phenomenon whereby an older system has a shorter remaining lifetime, in some statistical sense, that a newer one" (Bryson and Siddiqui [12]). Ageing notions of epoch times and interarrival times of NHPP have been studied extensively in the literature, see e.g. Baxter [7], Gupta and Kirmani[28] and Pellerey et al.[74]. The transmission of some ageing notions from a random sample to the corresponding order statistics have been investigated in Nagaraja [68], Misra et al. [64] and recently in Kundu et al. [47]. In Chapter 3, we study the preservation of some ageing notions from the underlying distribution functions to the SOS.

1.4.1. Definitions of stochastic orderings

Here, we give briefly a review of stochastic orders related to the *location*, the *magnitude*, the *dispersion* and the *shape* of random variables, respectively. Throughout, we shall use *increasing* to mean *non-decreasing* and *decreasing* to mean *non-increasing*. The following definitions introduce the stochastic orders that we will consider in this thesis.

Definition 1.4.1. Let X and Y be univariate random variables with cumulative distribution functions (cdf's) F and G, respectively. We say that X is smaller than Y in the *usual stochastic order* if $\overline{F}(t) \leq \overline{G}(t)$, for all t and in this case, we write $X \leq_{st} Y$ or $F \leq_{st} G$.

Recall from (1.1.1) the definition of the hazard rate function h_X of a random variable X. Let h_Y be the hazard rate function of another random variable Y.

Definition 1.4.2. X is said to be smaller than Y in the *hazard rate order*, denoted by $X \leq_{hr} Y$ or $F \leq_{hr} G$, if $h_X(t) \geq h_Y(t)$, for all t, or if $\overline{G}(t)/\overline{F}(t)$ is increasing in t for which the ratio is well defined.

Recall from (1.1.5) the definition of the reversed hazard rate function r_X of a random variable X. Let r_Y be the reversed hazard rate function of another random variable Y.

Definition 1.4.3. We say that X is smaller than Y in the *reversed hazard rate order* if $G(t)/F(t)$ is increasing in t for which the ratio is well defined, or if $r_X(t) \leq r_Y(t)$, for all t, denoted by $X \leq_{rh} Y$ or $F \leq_{rh} G$.

Definition 1.4.4. Let X and Y be univariate random variables with density functions (pdf's) f and g, respectively, such that $g(t)/f(t)$ is increasing in t for which the ratio is well defined. Then X is said to be smaller than Y in the *likelihood ratio order*, denoted by $X \leq_{lr} Y$ or $F \leq_{lr} G$.

The relationships among the four first orders are illustrated in the following diagram.

$$\begin{array}{ccc} X \leq_{lr} Y & \Rightarrow & X \leq_{hr} Y \\ \Downarrow & & \Downarrow \\ X \leq_{rh} Y & \Rightarrow & X \leq_{st} Y \end{array}$$

Next, we review the dispersive order that compare the *variability* or the *dispersion* of random variables.

Definition 1.4.5. We say that X is smaller than Y in the *dispersive order* if

$$F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha),$$

for all $0 < \alpha < \beta < 1$, where we write $X \leq_{\text{disp}} Y$ or $F \leq_{\text{disp}} G$.

MacGillivray and Blanda [55] defined distributions F and G to have the same shape if for some a and b, $F(t) = G(at + b)$, for all t. A basic concept to compare the *shape* in two probability distributions is through shape order as defined below. First, we recall the definition of a star-shaped function.

Definition 1.4.6. A function φ defined on $[0, \infty)$, which satisfies $\varphi(0) = 0$, is said to be *star-shaped* (anti star-shaped) if $\varphi(t)/t$ is increasing (decreasing) in t.

It should be noted that shape orders are preorders but not partial orders.

Definition 1.4.7. X is said to be smaller than Y in the *star order*, when the two random variables are non-negative, denoted by $X \leq_* Y$ or $F \leq_* G$, if $G^{-1}F(t)$ is star-shaped in t.

We shall also be using the concept of majorization in our discussion. Let $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$ denote the increasing arrangement of the components of the vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.

Definition 1.4.8. The vector $\boldsymbol{x}$ is said to be majorized by the vector $\boldsymbol{y}$, denoted by $\boldsymbol{x} \leq^m \boldsymbol{y}$, if

$$\sum_{i=1}^{j} x_{(i)} \geq \sum_{i=1}^{j} y_{(i)}, \quad \text{for } j = 1, \ldots, n-1 \quad \text{and} \quad \sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} y_{(i)}.$$

Functions that preserve the ordering of majorization are said to be Schur-convex, as one can see in the following definition.

Definition 1.4.9. A real valued function φ defined on a set $\mathcal{A} \in \Re^n$ is said to be *Schur-convex* (*Schur-concave*) on $\mathcal{A}$ if

$$\boldsymbol{x} \leq^m \boldsymbol{y} \text{ on } \mathcal{A} \Rightarrow \varphi(\boldsymbol{x}) \leq (\geq) \varphi(\boldsymbol{y}).$$

For extensive and comprehensive details on the theory of majorization orders and their applications, please refer to the excellent book of Marshall and Olkin [56].

1.4.2. Notions of ageing

Concepts of ageing describe how a component or a system improves or deteriorate with age. To be more specific, by ageing (respectively, antiageing) we mean a mathematical specification of degradation (respectively, upgradation) of a unit over

time. As we mentioned in Section 1.1, the software, usually, does not wear out, i.e, it does not ageing. However, the occurrence of software ageing in real systems has been documented in the literature, see e.g. Grottke et al. [27] and Matias and Freitas [58].

Many classes of life distributions can be categorized according to their ageing properties. Below, we define some of the most well known ageing notions.

Definition 1.4.10. Let X be a random variable with distribution function F and survival function $\overline{F}$. The random variable X (or its distribution) is said to be

i) increasing (decreasing) hazard rate or IHR (DHR) if $\overline{F}$ is logconcave (logconvex);

ii) increasing (decreasing) reversed hazard rate or IRHR (DRHR) if F is logconcave (logconvex);

iii) increasing (decreasing) hazard rate average or IHRA (DHRA) if $-\ln\overline{F}$ is star-shaped (anti star-shaped) when X is a non-negative random variable;

iv) increasing (decreasing) likelihood ratio or ILR (DLR) if f is logconcave (logconvex), when X is a continuous random variable.

It is well-known (see Marshall and Olkin [57]) that if X is a random variable with non-negative support then

$$\begin{array}{ccccc} \text{logconcave density} & \Rightarrow & \text{IHR} & \Rightarrow & \text{IHRA} \\ \Downarrow & & & & \\ \text{DRHR} & \Leftarrow & \text{concave distribution} & & \\ & & \Uparrow & & \\ \text{logconvex density} & \Rightarrow & \text{DHR} & \Rightarrow & \text{DHRA} \end{array}$$

1.5. SOFTWARE RELIABILITY MODELLING

The use of statistical methods in software engineering has been increasing in the last decades. In the context of this discipline, as we defined in Section 1.1, software reliability measures the probability that a piece of software runs without failing under certain operational conditions for a given time. In software testing, software is run under an operational profile, that is certain conditions simulating real usage and after a given test period, the software is modified in order to correct any observed faults. Testing then proceeds until the software is judged sufficiently reliable for release.

A software reliability model (SRM) is a mathematical tool to evaluate the software quantitatively. The SRM's have been extensively developed in the literature. Most SRM's are based on stochastic counting processes, such as binomial process, pure

birth process and nonhomogeneous Poisson process (NHPP). One may refer to two excellent books by Singpurwalla and Wilson [91] and Pham [77] on this topic. These stochastic models attempt to model either the times between successive failures of a piece of software or the number of failures in fixed time periods. Our classification scheme (see Figure 1.6) follows that of Singpurwalla and Wilson [91], and divides models into two types: Type I and Type II.

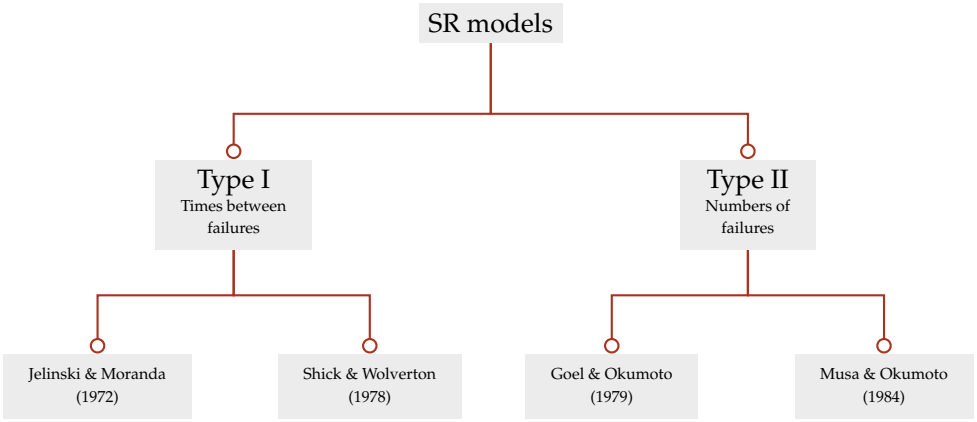


Figure 1.6: Classification of software reliability (SR) models

Type I models are those that model the times between successive failures. Under these types of models, the random variables $T_1, T_2, \ldots$, are modeled directly. This is often done by specifying the failure rate function for each random variable, h_i, $i = 1, 2, \ldots$, an then invoking the exponentiation formula (1.1.4) to obtain their survival function, $\overline{F}_i$. Typically, each h_i is a nondecreasing function on t, for $t \geq 0$, as we mentioned in Section 1.1, to reflect the fact that between failures the reliability of the software increases.

Type II models are those that model the number of failures up to a given time. These models are based on stochastic counting processes (see Section 1.2) for $N(t)$, the number of times the software fails in an interval $[0, t]$. The earliest and best known Type II models are those which assume that $N(t)$ is described by a Poisson process whose mean value function is based on assumptions about how the software experiences failure.

It is remarkable that a model of either type defines a model of the other. Specifically, for a sequence of interfailure times $T_1, T_2, \ldots$, for which a Type I model has been proposed, there is an implicit Type II model (cf. Kuo and Yang[49] and Singpurwalla and Wilson [91]), because

$$N(t) = \max\left\{ n | \sum_{i=1}^{n} T_i \leq t \right\},$$

and conversely, for a Type II model there is a Type I model, because with $T_0 = 0$, and $i = 2,3,\ldots,$

$$T_i = \inf\{t|N(t) = i\} - T_{i-1}.$$

It is noteworthy two differences between Type I and Type II models. First, the total number of potential failures of Type II models is assumed to be infinite, so that the number of observed failures is a random variable having a Poisson distribution, as opposed to a fixed number of faults N that is assumed by Type I models. Second, in the Type II models the interfailure times are dependent whereas in the Type I models they were typically assumed independent.

1.5.1. Type I models

The Type I group of models is used to study the program hazard rate per fault at the failure intervals. The hazard rate function of the i'th interfailure time of some of these models are reported in Table 1.1.

The first model to be widely known and used is the model by Jelinski and Moranda (JM) [33]. They assume that the software contains an unknown number, say N, of faults and that each time the software fails, a bug is detected and perfectly corrected. Furthermore, the failure rate of T_i is proportional to $N - i + 1$, the number of faults remaining in the code, that is, for some constant $\phi > 0$, the hazard rate at the i'th failure interval is given by

$$h_i(t) = \phi\,(N - i + 1)\,, \quad i = 1,\ldots,N. \tag{1.5.12}$$

The survival function is

$$\overline{F}_i(t) = e^{\phi(N-i+1)t}, \quad i = 1,\ldots,N.$$

The property of this model is that the failure rate is constant and the software during the testing stage is unchanged or frozen.

Table 1.1: Some Type I software reliability models

Jelinski-Moranda	Moranda
$h_i(t) = \phi\,(N - i + 1)$	$h_i(t) = Dk^{i-1}$
Goel-Okumoto	**Shick-Wolverton**
$h_i(t) = \phi\,(N - p(i-1))$	$h_i(t) = \phi\,(N - i - 1)t$

The Jelinski-Moranda model belongs to the class of exponential order statistic (EOS) model, defined in Section 1.3. The failures in a program will occur at times $X_1, X_2, \ldots$, measured from the beginning of debugging. This model assumed that all failures are similar, so the X_i's can be treated as identically distributed with density function

$$f(t) = \phi\, e^{-\phi t}, \quad t \geq 0.$$

Then, the order statistics $X_{1:n}, X_{2:n}, \ldots$ are the epoch times in which a failure of software takes place, and the interfailures are the spacings between the order statistics. Boland et al. in [10] demonstrated that for *any* independent random variables (not necessarily identically distributed), the order statistics are ordered with respect to the hazard rate ordering. Thus, the epoch times, S_i, are ordered in the hazard rate ordering, and from (1.5.12) it is easy to check that the interfailures, T_i, are also ordered according to the hazard rate ordering, since $h_i(t) = \phi\,(N-i+1) \geq h_{i+1}(t) = \phi\,(N-(i+1)+1)$.

A modification to the JM model is the *Geometric Model* developed by Moranda [66]. He proposed a new model in which the program failure rate function is initially a constant D and decreases geometrically at failure times. In this case, the hazard rate function of the i'th interfailure times is

$$h_i(t) = D k^i,$$

and its survival function is

$$\overline{F}_i(t) = e^{-tDk^i},$$

where $D > 0$ is the initial program failure rate and k is the parameter of a geometric function ($0 < k < 1$).

Goel and Okumoto [25] extend the JM model by assuming that a fault is removed with probability p whenever a failure occurs. This model is called the *JM model with imperfect debugging* and the hazard rate function of time between failures when the imperfect debugging is at the i'th failure interval becomes

$$h_i(t) = \phi\,(N - p(i-1))\,.$$

The survival function is

$$\overline{F}_i(t) = e^{-\phi(N-p(i-1))t}, \quad i = 1, \ldots, N.$$

The model by Jelinski and Moranda is a special case of the preceding when $p = 1$.

The model by Shick and Wolverton (SW) [86] is another modification of the JM model. They assumed that the hazard rate of T_i is proportional to both the number

of remaining faults in the software and the elapsed time since last failure. Thus, the hazard rate function between the $(i-1)$'th and the i'th failure can be expressed as

$$h_i(t) = \phi\,(N-i-1)\,t,$$

where ϕ and N are the same as that defined in the JM model.

1.5.2. Type II models

In this subsection we shall describe briefly some Type II software reliability models. The models described here are only a small subset of those which appear in the literature.

The Type II models provide another analytical framework for describing the software failure phenomenon during testing. Recall that in this case, we look at $N(t)$ as the number of failures to time t. Then, $N(t)$ is modeled by a Poisson distribution with mean $\Lambda(t)$, that is, $\mathbb{E}[N(t)] = \Lambda(t)$. Under such models the reliability of the software for a mission of duration t is simply $\Pr(N(t) = 0)$.

The Goel-Okumoto model [26], referred to hereafter as GO, is a NHPP variant of the JM model. The GO model assumes that the cumulative number of failures detected by time t is a NHPP and its expectation could be described by the mean value function

$$\Lambda(t) = a(1-e^{-bt}). \tag{1.5.13}$$

The intensity function is

$$\lambda(t) = \frac{d\Lambda(t)}{dt} = abe^{-bt}. \tag{1.5.14}$$

Observe that $\Delta(t) < \infty$ as $t \to \infty$. Therefore, this model cannot be applied to situations where new faults might be introduced in the process of debugging. Some NHPP models can incorporate the situation where new faults may be added during repairs, these models are the infinite failures models. It means that $\Delta(t) \to \infty$ as $t \to \infty$.

The Duane model [21], referred to hereafter as DU, originally devised for hardware reliability model, is an infinite failures model. This model is a NHPP with the expected number of failures

$$\Lambda(t) = at^b, \tag{1.5.15}$$

and the intensity function

$$\lambda(t) = abt^{b-1}. \tag{1.5.16}$$

This function is increasing for $b > 1$, decreasing for $b < 1$ and constant for $b = 1$. The DU model could be stochastically represented as a Weibull process, allowing for statistical procedures to be used in the application of this model in reliability growth.

In particular, this model is the counting process of the record values from a Weibull distribution.

In these NHPP models, usually parameter a represents the mean number of software failures that will eventually be detected, and parameter b represents the probability that a failure is detected in a constant period.

Musa and Okumoto [67] proposed another model for infinite failures. This NHPP is also called the *logarithm Poisson model*, referred to hereafter as MO. The mean value function is

$$\Lambda(t) = a\ln(1+bt), \quad t > 0, \tag{1.5.17}$$

and the intensity function is derived as

$$\lambda(t) = \frac{ab}{1+bt}. \tag{1.5.18}$$

Let us mention an homogeneous pure birth process, referred to hereafter as HPBP, for software reliability which is another variation of the JM model. This model, proposed by Boland and Singh [11], is a birth process approach to the geometric SRM (see Subsection 1.5.1). In this case, the cumulative number of failures detected by time t is a HPBP with birth rates

$$\lambda_i = D\cdot k^i, \quad i = 0,1,\ldots.$$

Boland and Singh [11] showed that the mean value function is

$$\Lambda(t) = Dt + \sum_{i=1}^{\infty}(-1)^i \frac{(Dt)^{i+1}}{(i+1)!}\prod_{j=1}^{i}(1-k^j), \tag{1.5.19}$$

and the intensity function

$$\lambda(t) = D\left(1+\sum_{i=1}^{\infty}(-1)^i \frac{(Dt)^i}{i!}\prod_{j=1}^{i}(1-k^j)\right). \tag{1.5.20}$$

Other types of mean value functions suggested by Ohba [72] and Yamada and Osaki [104], are the hyperexponential growth model and the Yamada-Osaki exponential growth model, respectively. Some of these models are reported in Table 1.2. We will review some well known Bayesian approaches to Type I and Type II models in Subsection 4.2.2. For more details on software reliability models, see e.g. Pham [77] and Singpurwalla and Wilson [91].

1.5.3. Stochastic comparisons in some well known SR models

One of the main aims of this thesis is to study stochastic properties of the times when a software failure occurs, S_i, and of the times between successive failures, T_i. Due

Table 1.2: Some Type II software reliability models

Goel-Okumoto	Musa-Okumoto
$\Lambda(t) = a\big(1-\exp(-bt)\big)$	$\Lambda(t) = a\ln(1+bt)$
Duane	**Boland and Singh**
$\Lambda(t) = at^b$	$\Lambda(t) = Dt + \sum_{i=1}^{\infty}(-1)^i \frac{(Dt)^{i+1}}{(i+1)!}\prod_{j=1}^{i}(1-k^j)$
Ohba	**Yamada-Osaki**
$\Lambda(t) = \sum_{i=1}^{n} a_i\big(1-\exp(-b_i t)\big)$	$\Lambda(t) = a\sum_{i=1}^{n} p_i\big(1-\exp(-b_i t)\big)$

to the relation between this random variables and the NHPP processes (see Section 1.2) we can apply the following results to Type II software reliability models.

Baxter [7] showed that the successive epoch times of a NHPP process are ordered in the hazard rate ordering.

Theorem 1.5.1 (Baxter [7]). *Let $\{N(t), t \geq 0\}$ be a nonhomogeneous Poisson process with mean function $\Lambda(t)$ (that is, $\Lambda(t) = E[N(t)]$, $t \geq 0$). Let $S_1, S_2, \ldots$ be the successive epoch times, then*

$$S_i \leq_{hr} S_{i+1}, \quad \text{for } i = 1, 2, \ldots.$$

Subsequently, Kochar [39] strengthened this result from the hazard rate ordering to the likelihood ratio ordering.

Theorem 1.5.2 (Kochar [39]). *Under the same assumptions as in Theorem 1.5.1, then*

$$S_i \leq_{lr} S_{i+1}, \quad \text{for } i = 1, 2, \ldots.$$

One may wonder whether the Theorems 1.5.1 and 1.5.2 can be strengthened from the epoch times to inter-epoch times of NHPP processes. Kochar [40] asserted that the inter-epoch times are ordered according to the hazard rate and the likelihood ratio orders under some conditions.

Theorem 1.5.3 (Kochar [40]). *Let $\{N(t), t \geq 0\}$ be a nonhomogeneous Poisson process with mean function $\Lambda(t)$. Let $T_1, T_2, \ldots$ be the successive inter-epoch times.*

i) If $\lambda(t)$ is decreasing (increasing) then $T_i \leq_{hr} (\geq_{hr}) T_{i+1}$, for $i = 1, 2, \ldots$.

ii) If $\lambda(t)e^{-\Lambda(t)}$ is log-convex (log-concave) then $T_i \leq_{lr} (\geq_{lr}) T_{i+1}$, for $i = 1, 2, \ldots$.

In the following examples we show applications of the previous theorems on some well known Type II software reliability models which we defined in Subsection 1.5.2.

Example 1.5.4 (**Goel-Okumoto SRM**). An application of Theorem 1.5.1 on the GO software reliability model can be seen in Figure 1.7. From (1.2.10), (1.5.13) and (1.5.14), it is easy to check that the hazard rate function of the i'th epoch time is

$$h_i(t) = \frac{abe^{-bt}\left(a\left(1-e^{-bt}\right)\right)^{i-1}}{(i-1)!\sum_{j=0}^{i-1}\frac{\left(a\left(1-e^{-bt}\right)\right)^{j}}{j!}}.$$

Note that $h_1(t) = \lambda(t)$.

According to (1.5.14), we know that $\lambda'(t) = -ab^2e^{-bt} \leq 0$, since $a,b > 0$, i.e., $\lambda(t)$ is decreasing. Thus, from Theorem 1.5.3(i), the times between failures of the GO model verify $T_i \leq_{hr} T_{i+1}$. On the other hand, $\lambda(t)e^{-\Lambda(t)}$ is log-convex if and only if $\left(\ln\lambda(t) - \Lambda(t)\right)$ is convex. It is easy to check that

$$\frac{d^2\left(\ln\lambda(t)-\Lambda(t)\right)}{dt^2} = -\lambda'(t) = ab^2e^{-bt} \geq 0,$$

then $T_i \leq_{lr} T_{i+1}$, for $i = 1,2,\ldots$, from Theorem 1.5.3(ii). ◁

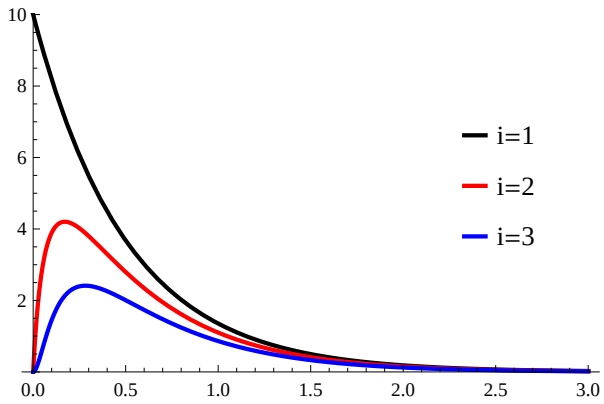


Figure 1.7: Hazard rate function of S_i of GO-SRM for $i = 1,2,3$ when $a = 5$ and $b = 2$

Example 1.5.5 (**Musa-Okumoto SRM**). It is immediately to prove that $T_i \leq_{hr} T_{i+1}$, since $\lambda'(t) = -\frac{ab^2}{(1+bt)^2} < 0$, from (1.5.18). And by (1.5.17), we get

$$\frac{d^2\left(\ln\lambda(t)-\Lambda(t)\right)}{dt^2} = \frac{(a+1)b^2}{(1+bt)^2} \geq 0,$$

then, again from Theorem 1.5.3(ii), $T_i \leq_{lr} T_{i+1}$, for $i = 1,2,\ldots$. ◁

Example 1.5.6 (**Duane SRM**). In this model, depending on the parameter b the times between failures are ordered in increasing or in decreasing hazard rate ordering. The intensity function $\lambda(t)$ is increasing for $b > 1$ and decreasing for $b < 1$, as we noted in Subsection 1.5.2. Thus, $T_i \geq_{hr} T_{i+1}$ when $b > 1$ and $T_i \leq_{hr} T_{i+1}$ when $b < 1$. In the same way, as

$$\frac{d^2\big(\ln\lambda(t) - \Lambda(t)\big)}{dt^2} = (1-b)\left(\frac{1}{t^2} + abt^{b-2}\right),$$

one can be proved that $T_i \geq_{lr} T_{i+1}$ when $b > 1$ and $T_i \leq_{lr} T_{i+1}$ when $b < 1$. ◁

1.6. STRUCTURE OF THIS THESIS

This thesis contains five chapters. Chapter 1 presents some basic definitions. After first providing a brief review of software reliability measures, it discusses the relationship between types of failure data and stochastic point processes used in the literature to model these kind of data, as well the relationship between the jumps of these stochastic point processes and models of ordered random variables. This is then followed by a brief review of some stochastic orderings in order to make comparisons of failure times or interfailure times based on their survival functions and hazard rate functions, among others. The chapter concludes with a review of some well known software reliability models.

The theoretical contributions of this thesis are developed in Chapters 2 and 3. The results of Chapter 2 concern stochastic orders for spacings of the order statistics of independent exponential random variables with different scale parameters. These results on stochastic orderings and spacings are motivated by the relationship between the spacings and the times elapsed between successive failures of a software program. Due to the complicated expression of the distribution in the non-iid case, only limited results are found in the literature. In Torrado et al. [97], we investigate the hazard rate ordering of simple spacings and normalized spacings of a sample of heterogeneous exponential random variables. These results can be found in the first part of Chapter 2. In the second part of this chapter, we study the two sample problem. Specifically, we compare both simple spacings and normalized spacings from two samples of heterogeneous exponential random variables according to the likelihood ratio ordering. We also show applications of these results to multiple-outlier models.

In Chapter 3, motivated by the equality in distribution between SOS and the first n epoch times of a NHPB process, we consider the problem of comparing the components of sequential k-out-of-n systems according to magnitude and location orders. Distributional and stochastic properties of the lifetimes of sequential k-out-of-n sys-

tems have been studied for some researchers, such as Cramer and Kamps [18], Zhuang and Hu [105], Navarro and Burkschat [69] and recently for Torrado et al. [96]. In particular, this chapter discusses conditions on the underlying distribution functions on which the SOS are based, to obtain ageing notions and stochastic comparisons of SOS. We also present a NHPB process approach to software reliability modelling.

Following a Bayesian analysis approach, we develop in Chapter 4 a new procedure to predict both interfailure times and numbers of software failures using metrics information. Research on how to predict software failures accurately is of great practical importance. A large number of models have been proposed to address this topic, but a few incorporate some significant metrics data observed in software testing. In particular, we consider the prediction of times between software failures or numbers of failures in a given time, when it is assumed that the software is possibly imperfectly repaired after each failure and, that software metrics information for each version of the software is available. Our model is a hierarchical non-parametric regression model based on exponential interfailure times or Poisson failure counts where the rates are modeled as Gaussian processes with software metrics data such as lines of code, complexity measures or even computer execution times are used as inputs. We undertake fully Bayesian inference based on MCMC and illustrate our approach with real software failure data.

In Chapter 5 we show some general conclusions and describe the most significant contributions of this thesis.

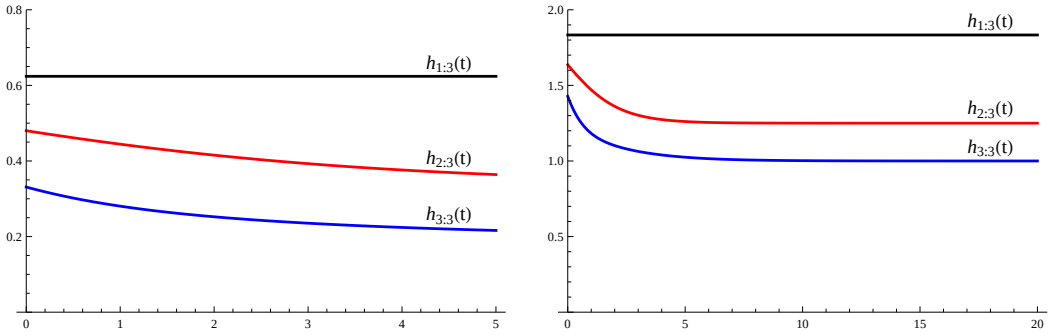
CHAPTER 2

Spacings based on order statistics

Spacings and their functions are important in statistics, in general, and in particular in the context of life testing and reliability models. In these context, as we mentioned in the introductory chapter (see section 1.3), an n component system that works if and only if at least k of the n components work is called a k-out-of-n system. The lifetime of a k-out-of-n system is usually described by the $(n-k+1)$'th order statistic from a random sample $X_1, X_2, \ldots, X_n$ where the variable X_i represents the lifetime or failure time of the i'th component of the system, $1 \leq i \leq n$. The times between failures of components in a k-out-of-n system correspond with the spacings associated with order statistics. In the conventional modelling of these structures, the component lifetimes are supposed to be independent and identically distributed random variables. We will see in this and in the next chapter as some of these constrains can be relaxed.

A lot of work has been done in the literature on different aspects of order statistics and spacings, see [3, 4] for a review. However most of this work has been confined to the case when the observations are iid, but in many practical situations, like in reliability theory, the observations are not necessarily iid. For example, in software reliability, failure times of a software program are modeled as order statistics of independent nonidentically distributed (inid) exponential random variables. According to Miller [63], these models are called EOS, as we mentioned in Section 1.3. The JM, GO and MO models are all special cases of EOS models. It is well known that OS from heterogeneous exponential random variables are ordered with respect to the hazard rate ordering, i.e., the failure times, $S_1, \ldots, S_n$, of EOS-SR models verify that $S_i \leq_{hr} S_{i+1}$, for $i = 1, \ldots, n-1$. Thus, a natural question to ask is whether the spacings from expo-

nential random variables with different scale parameters are also ordered according to the hazard rate ordering. In Figure 2.1, we show two examples on this, when $\lambda_i = a\,b^i$, $a > 0$, $0 < b < 1$ and when $\lambda_i = a\,i^{-b}$, $a > 0$, $1 < b < \infty$, which are case 3 (geometric rates) and case 4 (power rates) in Miller [63], respectively.



(a) $\lambda_i = a\,b^i$, $a = 3$, $b = 0.4$ (b) $\lambda_i = a\,i^{-b}$, $a = 3$, $b = 1.1$

Figure 2.1: Hazard rate function of spacings for two EOS software reliability models

Specifically, Figure 2.1a and Figure 2.1b present the hazard rate function, $h_{i:3}(t)$, of normalized spacings from three heterogeneous exponential random variables having hazard rate $\lambda_i = a\,b^i$, $a = 3$, $b = 0.4$ and $\lambda_i = a\,i^{-b}$, $a = 3$, $b = 1.1$, respectively. As seen from these figures, the normalized spacings are ordered according to the hazard rate ordering in both cases.

Let $X_1, X_2, \ldots, X_n$ be independent, but not necessarily identically distributed, random variables. Recall from Section 1.3 that

$$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$$

are the order statistics of these random variables and,

$$D_{i:n} = X_{i:n} - X_{i-1:n} \quad \text{and} \quad D^*_{i:n} = (n-i+1)\,(X_{i:n} - X_{i-1:n})\,,$$

for $i = 1, \ldots, n$, with $X_{0:n} \equiv 0$, are respectively called *simple spacings* and *normalized spacings*.

Many authors have studied the stochastic properties of spacings from independent and identically random variables. In this context, the exponential distribution possesses a lack of ageing property, so that, the failure rate is constant and the spacings correspond to times elapsed between successive failures of components in a system. The following remarkable property of the exponential distribution was first demonstrated by Sukhatme [95] in 1937.

Theorem 2.0.1. *If $X_1, X_2, \ldots, X_n$ are independent and identically exponential random variables, then $D_{1:n}, \ldots, D_{n,n}$ are iid random variables having the same exponential distribution.*

Such a characterization may not hold for other distributions or when the observations are not independent and identically distributed.

The objective of this chapter is first to discuss some recent results on stochastic comparisons between spacings of heterogeneous samples and present some extensions. Specifically, we study stochastic orderings between the second an the third spacings in the one sample problem, and also, we obtain results in the two sample problem. In this case, we show some applications to a multiple-outlier model.

The chapter is organized as follows. In Section 2.1, we introduce the probability density function (pdf) of normalized spacings, and give two useful lemmas which will be used in the following sections. We investigate, in Section 2.2, the hazard rate ordering of simple spacings and normalized spacings of a sample from heterogeneous exponential random variables. Section 2.3 is devoted to stochastic comparisons of both simple and normalized spacings in two sample problem. Finally, conclusions and possible extensions to this work are considered in Section 2.4.

2.1. PRELIMINARY RESULTS

For heterogeneous but independent exponential random variables, Kochar and Korwar [43] (hereafter K&K) proved that, for $i \in \{2, \ldots, n\}$, the distribution of D_i^* is a mixture of independent exponential random variables with p.d.f.:

$$f_i(t) = \sum_{\mathbf{r}_n} \frac{\prod_{k=1}^{n} \lambda_k}{\prod_{k=1}^{n} \left(\sum_{j=k}^{n} \lambda(r_j) \right)} \cdot \left(\frac{\sum_{j=i}^{n} \lambda(r_j)}{n-i+1} \right) \cdot \exp\left\{ \frac{-t \sum_{j=i}^{n} \lambda(r_j)}{n-i+1} \right\}, \tag{2.1.1}$$

where $\mathbf{r}_n = (r_1, \ldots, r_n)$ is a permutation of $(1, \ldots, n)$ and $\lambda(i) = \lambda_i$. They also showed that $D_{1:n}^*$ is independent of $(D_{2:n}^*, \ldots, D_{n:n}^*)$.

Observe that in (2.1.1), the term $\sum_{j=i}^{n} \lambda(r_j)$ coincides for all permutations $\mathbf{r}_n$ that have the same groups of λ_k's in the last $n-i+1$ positions. This remark permits us to simplify the notation as follows. Let

$$\beta_{m_j}^{(i)} = \frac{\sum_{\ell=i}^{n} \lambda(r_\ell)}{n-i+1}, \tag{2.1.2}$$

where m_j indicates a group of indices of size $n-i+1$. Then, (2.1.1) can be written as

$$f_i(t) = \sum_{j=1}^{M_i} \Delta(\beta_{m_j}^{(i)}, n)\, \beta_{m_j}^{(i)}\, e^{-t\beta_{m_j}^{(i)}}, \tag{2.1.3}$$

where $M_i = \binom{n}{n-i+1}$ and

$$\Delta(\beta_{m_j}^{(i)}, n) = \sum_{\mathbf{r}_{i-1,m_j}} \left(\prod_{k \in H_{m_j}} \lambda_k \right) \left[\prod_{\ell=1}^{i-1} \left\{ \sum_{\substack{u=\ell \\ r_u \in H_{m_j}}}^{i-1} \lambda_{r_u} + (n-i+1)\beta_{m_j}^{(i)} \right\} \right]^{-1}, \tag{2.1.4}$$

where $H_{m_j} = \{1, \ldots, n\} - m_j$ and the outer summation is being taken over all permutations of the elements of H_{m_j}. Note that (2.1.4) and Equation 2.3 of K&K are equivalent, although with different notation.

Before proceeding to our main results, let us first recall two lemmas, which will be used repeatedly in the following sections.

Lemma 2.1.1 (Lemma 3.1., in K&K [43]). *Let $\Delta(\beta_{m_j}^{(i)}, n)$ be as defined in (2.1.4). Suppose that m_1 and m_2 are two subsets of $\{1, \ldots, n\}$ of size $n-i+1$ $(1 < i \leq n)$ and that they have all but one element in common. Denote the different element in m_1 by a_1 and that in m_2 by a_2. Then*

$$\lambda_{a_1} \Delta(\beta_{m_1}^{(i)}, n) \geq \lambda_{a_2} \Delta(\beta_{m_2}^{(i)}, n), \qquad \text{if} \quad \lambda_{a_2} \geq \lambda_{a_1}.$$

Lemma 2.1.2 (Chebyshev's inequality, Theorem 1, in Mitrinovic [65]). *Let $a_1 \leq a_2 \leq \ldots \leq a_n$ and $b_1 \leq b_2 \leq \ldots \leq b_n$ be two increasing sequences of real numbers. Then*

$$n \sum_{i=1}^{n} a_i b_i \geq \left(\sum_{i=1}^{n} a_i \right) \left(\sum_{i=1}^{n} b_i \right).$$

2.2. THE ONE SAMPLE PROBLEM

Many authors have studied the stochastic properties of spacings from restricted families of distributions. If $X_1, X_2, \ldots, X_n$ is a random sample from a *decreasing hazard rate* (DHR) distribution (see Definition 1.4.10), then it has been proved by Barlow and Proschan [5] that the successive normalized spacings are stochastically increasing. Kochar and Kirmani [42] strengthened this result from stochastic ordering to hazard rate ordering, that is, for $i = 1, \ldots, n-1$

$$D_{i:n}^* \leq_{hr} D_{i+1:n}^*. \tag{2.2.5}$$

The corresponding problem when the random variables are not identically distributed has also been well studied. In particular, Pledger and Proschan [78] proved that if the scale parameters of the exponential distributions are not all equal then the i'th normalized spacing is stochastically smaller than the $(i+1)$'th normalized spacing.

Theorem 2.2.1 (Pledger and Proschan [78]). *If $X_1, X_2, \ldots, X_n$ are independent exponential random variables with X_i having hazard rate λ_i , $i = 1, \ldots, n$, then*

$$D^*_{i:n} \leq_{st} D^*_{i+1:n},$$

for $i = 1, \ldots, n-1$.

K&K [43] conjectured that a result similar to (2.2.5) holds in the case when X_1, $X_2, \ldots, X_n$ are independent exponential random variables with possibly unequal scale parameters. They proved this conjecture for $n = 3$, and also obtained the following result on likelihood ratio ordering between $D^*_{1:n}$ and $D^*_{i:n}$ for $1 < i \leq n$.

Theorem 2.2.2 (K&K [43]). *Let $X_1, X_2, \ldots, X_n$ be independent exponential random variables with X_i having hazard rate λ_i , $i = 1, \ldots, n$. Then*

$$D^*_{1:n} \leq_{lr} D^*_{i:n},$$

for $i = 2, \ldots, n$.

Khaledi and Kochar [37] demostrated the conjecture of K&K in the case that the random variables $X_1, \ldots, X_n$ follow a single-outlier model with parameters λ and λ^*, that is, when $\lambda_1 = \cdots = \lambda_{n-1} = \lambda$ and $\lambda_n = \lambda^*$.

Theorem 2.2.3 (Khaledi and Kochar [37]). *Let $X_1, X_2, \ldots, X_n$ follow the single-outlier exponential model with parameters λ and λ^*. Then*

$$D^*_{i:n}(n-1, 1) \leq_{hr} D^*_{i+1:n}(n-1, 1),$$

for $i = 1, \ldots, n-1$.

The reader is referred to Khaledi and Kochar [38] for a review of some further results in the area of stochastic comparisons of order statistics and spacings.

In a multiple-outlier exponential model, Wen et al.[99] established likelihood ratio ordering between consecutive simple spacings. In this case, $X_1, X_2, \ldots, X_p$ is a random sample of size p from an exponential distribution with failure rate λ and $X_{p+1}, \ldots, X_n$ is another independent random sample of size q from an exponential distribution with failure rate λ^* where $q = n - p \geq 1$, $p \geq 1$.

Theorem 2.2.4 (Wen et al.[99]). *Let $X_1, X_2, \ldots, X_n$ follow the multiple-outlier exponential model with parameters λ and λ^*. Then*

$$D_{i:n}(p, q) \leq_{lr} D_{i+1:n}(p, q),$$

for $p \geq 1$, $q \geq 1$ and $i = 1, \ldots, n-1$, where $D_{i:n}(p, q)$ is the i'th simple spacing of $X_1, X_2, \ldots, X_n$.

In a recent paper, Hu et al.[31] established a result similar to Theorem 2.2.2 for the first and the second simple spacings of nonidentical independent exponential random variables.

Theorem 2.2.5 (Hu et al.[31]). *Let $X_1, X_2, \ldots, X_n$ be independent exponential random variables with X_i having hazard rate λ_i, $i = 1, \ldots, n$. Then*

$$D_{1:n} \leq_{lr} D_{2:n}.$$

Furthermore, they proved that, if $\lambda_i + \lambda_j \geq \lambda_k$ for all distinct i, j and k, then

$$D_{n-1:n} \leq_{lr} D_{n:n}.$$

The purpose of this section is to investigate the hazard rate ordering of simple spacings and normalized spacings of heterogeneous exponential random variables. We present some new results on hazard rate ordering concerning the second and the third simple spacings and normalized spacings of heterogeneous exponential random variables.

Observing (2.1.3), note that, by definition, $D^*_{i:n} \leq_{hr} D^*_{i+1:n}$ if and only if

$$h_i(t) = \frac{\sum_{j=1}^{M_i} \Delta(\beta^{(i)}_{m_j}, n)\, \beta^{(i)}_{m_j}\, e^{-t\beta^{(i)}_{m_j}}}{\sum_{j=1}^{M_i} \Delta(\beta^{(i)}_{m_j}, n)\, e^{-t\beta^{(i)}_{m_j}}} \geq \frac{\sum_{j=1}^{M_{i+1}} \Delta(\beta^{(i+1)}_{m_j}, n)\, \beta^{(i+1)}_{m_j}\, e^{-t\beta^{(i+1)}_{m_j}}}{\sum_{j=1}^{M_{i+1}} \Delta(\beta^{(i+1)}_{m_j}, n)\, e^{-t\beta^{(i+1)}_{m_j}}} = h_{i+1}(t),$$

which can be rewritten as

$$\sum_{j=1}^{M_{i+1}} \sum_{k=1}^{M_i} \Delta(\beta^{(i)}_{m_k}, n)\Delta(\beta^{(i+1)}_{m_j}, n)\, e^{-t\left(\beta^{(i)}_{m_k} + \beta^{(i+1)}_{m_j}\right)} \left(\beta^{(i)}_{m_k} - \beta^{(i+1)}_{m_j}\right) \geq 0. \tag{2.2.6}$$

Throughout this chapter we shall suppose, without loss of generality, that the λ_i's are in increasing order.

Next, we give an useful lemma which will be used later.

Lemma 2.2.6. *Let $\beta^{(i)}_{m_k}$ be as defined in (2.1.2), then*

$$\sum_{k=1}^{M_i} \sum_{j=1}^{M_{i+1}} \left(\beta^{(i)}_{m_k} - \beta^{(i+1)}_{m_j}\right) = 0.$$

PROOF.

$$\begin{aligned}\sum_{j=1}^{M_{i+1}}\sum_{k=1}^{M_i}\left(\beta_{m_k}^{(i)}-\beta_{m_j}^{(i+1)}\right) &= \sum_{k=1}^{M_i} M_{i+1}\,\beta_{m_k}^{(i)} - \sum_{j=1}^{M_{i+1}} M_i\,\beta_{m_j}^{(i+1)} \\ &= \sum_{\ell=1}^{n}\binom{n}{n-i}\binom{n-1}{n-i}\frac{\lambda_\ell}{n-i+1} - \sum_{\ell=1}^{n}\binom{n}{n-i+1}\binom{n-1}{n-i-1}\frac{\lambda_\ell}{n-i} \\ &= \left[\binom{n}{n-i}\binom{n-1}{n-i}\frac{1}{n-i+1} - \binom{n}{n-i+1}\binom{n-1}{n-i-1}\frac{1}{n-i}\right]\sum_{\ell=1}^{n}\lambda_\ell \\ &= 0\,,\end{aligned}$$

since

$$\binom{n}{n-i}\binom{n-1}{n-i}\frac{1}{n-i+1} = \binom{n}{n-i+1}\binom{n-1}{n-i-1}\frac{1}{n-i}\,. \qquad \square$$

2.2.1. Normalized spacings

It is an open problem in the literature, as K&K pointed out in [43], whether successive normalized spacings are ordered in the hazard rate ordering. We will partially solve this problem here.

Our first new result shows that, in general, the second normalized spacing is smaller than the third normalized spacing according to the hazard rate ordering.

Theorem 2.2.7. *Let $X_1,\ldots,X_n$ be independent exponential random variables such that X_i has hazard rate λ_i, for $i=1,\ldots,n$, then*

$$D^*_{2:n} \leq_{hr} D^*_{3:n}, \quad \text{for all } n.$$

PROOF. Observing equation (2.2.6), we have to show

$$\sum_{j=1}^{M_3}\sum_{k=1}^{M_2}\Delta(\beta_{m_k}^{(2)},n)\Delta(\beta_{m_j}^{(3)},n)\,e^{-t\left(\beta_{m_k}^{(2)}+\beta_{m_j}^{(3)}\right)}\left(\beta_{m_k}^{(2)}-\beta_{m_j}^{(3)}\right)\geq 0. \tag{2.2.7}$$

To do this, we consider those values of $\beta_{m_k}^{(2)}-\beta_{m_j}^{(3)}$ which add zero for each $k=1,\ldots,M_2$ and $j=1,\ldots,M_3$. To illustrate this idea, we show in Figure 2.2 the representation of the structure of the matrix of $\beta_{m_k}^{(2)}-\beta_{m_j}^{(3)}$ up to $n=6$. To simplify the notation in this particular case ($i=2$), we define

$$\beta_{(j,k)}^{(3)} = \frac{1}{n-2}\sum_{\substack{h=1\\ h\notin\{j,k\}}}^{n}\lambda_h \quad \text{and} \quad \beta_u^{(2)} = \frac{1}{n-1}\sum_{\substack{h=1\\ h\neq u}}^{n}\lambda_h, \quad \text{for } u=1,\ldots,M_2.$$

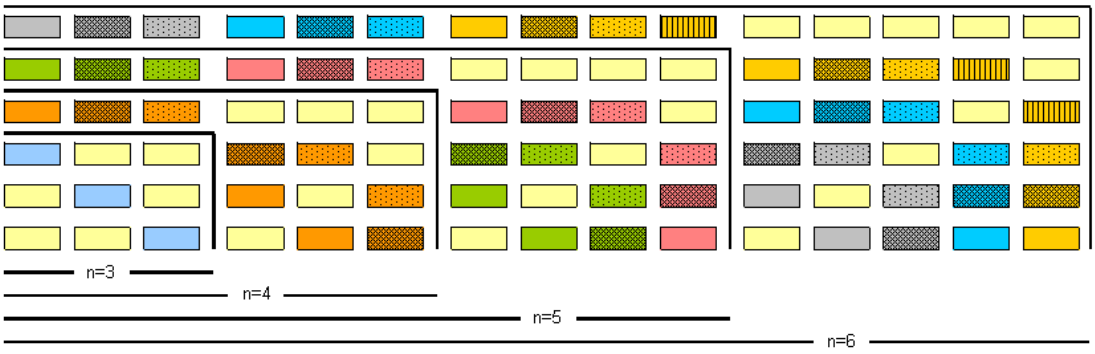


Figure 2.2: Representation of the matrix of $\beta^{(2)}_{m_k} - \beta^{(3)}_{m_j}$ up to $n = 6$

There are $n-1$ elements in each row of these matrices that sum to zero, that is,

$$\sum_{\substack{k=1\\k\neq u}}^{n} \left(\beta^{(2)}_u - \beta^{(3)}_{(u,k)}\right) = (n-1)\beta^{(2)}_u - \sum_{\substack{k=1\\k\neq u}}^{n} \beta^{(3)}_{(u,k)} = 0. \tag{2.2.8}$$

These values correspond to the yellow squares in the Figure 2.2. We are interested in proving that

$$C_u = \Delta(\beta^{(2)}_u, n)\, e^{-t\beta^{(2)}_u} \sum_{\substack{k=1\\k\neq u}}^{n} \Delta(\beta^{(3)}_{(u,k)}, n)\, e^{-t\beta^{(3)}_{(u,k)}} \left(\beta^{(2)}_u - \beta^{(3)}_{(u,k)}\right) \geq 0, \tag{2.2.9}$$

for $u = 1, \ldots, M_2$.

Notice that $a_{u,k} = \beta^{(2)}_u - \beta^{(3)}_{(u,k)}$ and $\exp\left\{-t\,\beta^{(3)}_{(u,k)}\right\}$ are two sequences increasing in $k \in \{1,\ldots,n\} - u$. It follows from Lemma 2.1.1

$$\Delta(\beta^{(3)}_{(u,k)}, n) \geq \Delta(\beta^{(3)}_{(u,k')}, n) \quad \text{for} \quad k > k'.$$

Then $b_{u,k} = \Delta(\beta^{(3)}_{(u,k)}, n)\, e^{-t\beta^{(3)}_{(u,k)}}$ are increasing in k. Finally, by Lemma 2.1.2, we conclude that

$$C_u \geq \frac{1}{n-1}\Delta(\beta^{(2)}_u, n)\, e^{-t\beta^{(2)}_u} \left(\sum_{\substack{k=1\\k\neq u}}^{n} \Delta(\beta^{(3)}_{(u,k)}, n)\, e^{-t\beta^{(3)}_{(u,k)}}\right) \left(\sum_{\substack{k=1\\k\neq u}}^{n} \left(\beta^{(2)}_u - \beta^{(3)}_{(u,k)}\right)\right) = 0,$$

since (2.2.8) holds.

We group the remaining values of $\beta^{(2)}_u - \beta^{(3)}_{(u,k)}$ in $\binom{n}{3}$ diagonals, each of them has a different color and aspect in the Figure 2.2. We fix a combination of three elements $j < k < \ell$ so that

$$a_{u,1} = \beta^{(2)}_j - \beta^{(3)}_{(k,\ell)} \geq a_{u,2} = \beta^{(2)}_k - \beta^{(3)}_{(j,\ell)} \geq a_{u,3} = \beta^{(2)}_\ell - \beta^{(3)}_{(j,k)}.$$

Since $\Delta(\beta_j^{(2)},n)=\frac{\lambda_j}{s_n}$ where $s_n=\sum_{h=1}^{n}\lambda_h$, we have from Lemma 2.1.1 that

$$\Delta(\beta_j^{(2)},n)\Delta(\beta_{(k,\ell)}^{(3)},n)\geq\Delta(\beta_k^{(2)},n)\Delta(\beta_{(j,\ell)}^{(3)},n)\geq\Delta(\beta_\ell^{(2)},n)\Delta(\beta_{(j,k)}^{(3)},n). \tag{2.2.10}$$

Then, again from Lemma 2.1.2

$$\sum_{h=1}^{3}a_{u,h}\,b_{u,h}\geq\frac{1}{3}\left(\sum_{h=1}^{3}a_{u,h}\right)\left(\sum_{h=1}^{3}b_{u,h}\right),\quad\text{for}\quad u=1,\ldots,\binom{n}{3},$$

where

$$\begin{aligned}
b_{u,1}&=\Delta(\beta_j^{(2)},n)\Delta(\beta_{(k,\ell)}^{(3)},n)e^{-t\left(\beta_j^{(2)}+\beta_{(k,\ell)}^{(3)}\right)},\\
b_{u,2}&=\Delta(\beta_k^{(2)},n)\Delta(\beta_{(j,\ell)}^{(3)},n)e^{-t\left(\beta_k^{(2)}+\beta_{(j,\ell)}^{(3)}\right)},\\
b_{u,3}&=\Delta(\beta_\ell^{(2)},n)\Delta(\beta_{(j,k)}^{(3)},n)e^{-t\left(\beta_\ell^{(2)}+\beta_{(j,k)}^{(3)}\right)}.
\end{aligned}$$

Let $A_u=\sum_{h=1}^{3}a_{u,h}$ and $B_u=\sum_{h=1}^{3}b_{u,h}$ be. Now, we group diagonals so that,

$$\sum_{u\in\text{group }1}A_u\geq\sum_{u'\in\text{group }2}A_{u'}.$$

Each group is formed by the (three or more) diagonals in Figure 2.2 which have the same color. Then, it is necessary to prove that the respective B_u are also ordered. One can see from Lemma 2.1.1 that

$$\sum_{u\in\text{group }1}B_u\geq\sum_{u'\in\text{group }2}B_{u'},$$

holds. Thus, from Lemma 2.1.2, we have

$$\sum_{u\in\text{group }1}A_uB_u+\sum_{u'\in\text{group }2}A_{u'}B_{u'}\geq\left(\sum_{u\in\text{group }1}A_u+\sum_{u'\in\text{group }2}A_{u'}\right)\left(\sum_{u\in\text{group }1}B_u+\sum_{u'\in\text{group }2}B_{u'}\right).$$

In this way, we can apply Lemmas 2.1.1 and 2.1.2 as many times as necessary until we can obtain that

$$\sum_{u=1}^{\binom{n}{3}}\sum_{h=1}^{3}a_{u,h}\,b_{u,h}\geq\left(\sum_{u=1}^{\binom{n}{3}}\sum_{h=1}^{3}a_{u,h}\right)\left(\sum_{u=1}^{\binom{n}{3}}\sum_{h=1}^{3}b_{u,h}\right)=0,$$

since, by Lemma 2.2.6, the sums of the differences of the betas is equal to zero. □

For a better understanding of the previous proof we recommend studying the proof for $n=4$ located in Appendix 2.5.

We prove below that, for $n=4$, the successive normalized spacings from heterogeneous exponential random variables are increasing in hazard rate ordering, that is,

$$D^*_{1:4} \leq_{hr} D^*_{2:4} \leq_{hr} D^*_{3:4} \leq_{hr} D^*_{4:4}.$$

K&K established likelihood ratio ordering between the first normalized spacing and the others, in particular $D^*_{1:n} \leq_{hr} D^*_{2:n}$. And by Theorem 2.2.7 we know that, for any n, $D^*_{2:n} \leq_{hr} D^*_{3:n}$, so we have to show $D^*_{3:4} \leq_{hr} D^*_{4:4}$.

Theorem 2.2.8. *Under the same assumptions as those in Theorem 2.2.7, then*

$$D^*_{3:4} \leq_{hr} D^*_{4:4}.$$

PROOF. Considering $i=3$ in (2.2.6), we have to show that

$$\sum_{j=1}^{M_4}\sum_{k=1}^{M_3} \Delta(\beta^{(3)}_{m_k},4)\Delta(\beta^{(4)}_{m_j},4)\, e^{-t\left(\beta^{(3)}_{m_k}+\beta^{(4)}_{m_j}\right)} \left(\beta^{(3)}_{m_k}-\beta^{(4)}_{m_j}\right) \geq 0. \tag{2.2.11}$$

First, examine the values of $\beta^{(3)}_{m_k}-\beta^{(4)}_{m_j}$ where $M_3=6$ and $M_4=4$.

$$\begin{pmatrix}
\frac{\lambda_3+\lambda_4}{2}-\lambda_1 & \frac{\lambda_3+\lambda_4}{2}-\lambda_2 & \frac{\lambda_3+\lambda_4}{2}-\lambda_3 & \frac{\lambda_3+\lambda_4}{2}-\lambda_4 \\
\frac{\lambda_2+\lambda_4}{2}-\lambda_1 & \frac{\lambda_2+\lambda_4}{2}-\lambda_2 & \frac{\lambda_2+\lambda_4}{2}-\lambda_3 & \frac{\lambda_2+\lambda_4}{2}-\lambda_4 \\
\frac{\lambda_2+\lambda_3}{2}-\lambda_1 & \frac{\lambda_2+\lambda_3}{2}-\lambda_2 & \frac{\lambda_2+\lambda_3}{2}-\lambda_3 & \frac{\lambda_2+\lambda_3}{2}-\lambda_4 \\
\frac{\lambda_1+\lambda_4}{2}-\lambda_1 & \frac{\lambda_1+\lambda_4}{2}-\lambda_2 & \frac{\lambda_1+\lambda_4}{2}-\lambda_3 & \frac{\lambda_1+\lambda_4}{2}-\lambda_4 \\
\frac{\lambda_1+\lambda_3}{2}-\lambda_1 & \frac{\lambda_1+\lambda_3}{2}-\lambda_2 & \frac{\lambda_1+\lambda_3}{2}-\lambda_3 & \frac{\lambda_1+\lambda_3}{2}-\lambda_4 \\
\frac{\lambda_1+\lambda_2}{2}-\lambda_1 & \frac{\lambda_1+\lambda_2}{2}-\lambda_2 & \frac{\lambda_1+\lambda_2}{2}-\lambda_3 & \frac{\lambda_1+\lambda_2}{2}-\lambda_4
\end{pmatrix} \tag{2.2.12}$$

This construction was motivated in the proof of Theorem 3.6 of K&K. Our main idea is to find coefficients of the matrix (2.2.12) which sum to zero. The groups of coefficients can be divided into two types:

$$\left(\frac{\lambda_j+\lambda_k}{2}-\lambda_j\right)+\left(\frac{\lambda_j+\lambda_k}{2}-\lambda_k\right)=0,$$

$$\left(\frac{\lambda_k+\lambda_\ell}{2}-\lambda_j\right)+\left(\frac{\lambda_j+\lambda_\ell}{2}-\lambda_k\right)+\left(\frac{\lambda_j+\lambda_k}{2}-\lambda_\ell\right)=0,$$

for $j,k,\ell=1,\ldots,4$ and $j<k<\ell$. To simplify the notation let $\beta^{(3)}_{m_k}=\frac{\lambda_j+\lambda_\ell}{2}$ and $\beta^{(4)}_{m_j}=\lambda_j$ be, if $m_k=(j,\ell)$ and $m_j=j$, respectively. Then, (2.2.11) can be written as

$$\sum_{j=1}^{4}\sum_{k=j+1}^{4} A_{(j,k)}+\sum_{u=1}^{4} B_u \geq 0,$$

where

$$A_{(j,k)} = \Delta(\beta^{(3)}_{(j,k)},4)\, e^{-t\left(\frac{\lambda_j+\lambda_k}{2}\right)} \times \left[\Delta(\beta^{(4)}_j,4)\, e^{-t\lambda_j}\left(\frac{\lambda_j+\lambda_k}{2}-\lambda_j\right) + \Delta(\beta^{(4)}_k,4)\, e^{-t\lambda_k}\left(\frac{\lambda_j+\lambda_k}{2}-\lambda_k\right)\right],$$

and

$$\begin{aligned} B_u =& \Delta(\beta^{(3)}_{(k,\ell)},4)\Delta(\beta^{(4)}_j,4)\, e^{-t\left(\frac{\lambda_k+\lambda_\ell}{2}+\lambda_j\right)}\left(\frac{\lambda_k+\lambda_\ell}{2}-\lambda_j\right) \\ &+\Delta(\beta^{(3)}_{(j,\ell)},4)\Delta(\beta^{(4)}_k,4)\, e^{-t\left(\frac{\lambda_j+\lambda_\ell}{2}+\lambda_k\right)}\left(\frac{\lambda_j+\lambda_\ell}{2}-\lambda_k\right) \\ &+\Delta(\beta^{(3)}_{(j,k)},4)\Delta(\beta^{(4)}_\ell,4)\, e^{-t\left(\frac{\lambda_j+\lambda_k}{2}+\lambda_\ell\right)}\left(\frac{\lambda_j+\lambda_k}{2}-\lambda_\ell\right), \end{aligned}$$

where $u \notin \{j,k,\ell\}$. We divide the proof into two parts according to different types of addition. First, we will show that $A_{(j,k)}$ are positive for all $j<k$. After some manipulations we can see that

$$A_{(j,k)} = \Delta(\beta^{(3)}_{(j,k)},4)\, e^{-t\left(\frac{\lambda_j+\lambda_k}{2}\right)}\left(\frac{\lambda_k-\lambda_j}{2}\right)\left[\Delta(\beta^{(4)}_j,4)\, e^{-t\lambda_j} - \Delta(\beta^{(4)}_k,4)\, e^{-t\lambda_k}\right].$$

Lemma 2.1.1 and $\lambda_j \le \lambda_k$ imply that

$$1 \le \frac{\lambda_k}{\lambda_j} \le \frac{\Delta(\beta^{(4)}_j,4)}{\Delta(\beta^{(4)}_k,4)}.$$

Therefore $A_{(j,k)} \ge 0$ since $\Delta(\beta^{(4)}_j,4) \ge \Delta(\beta^{(4)}_k,4)$ and $e^{-t\lambda_j} \ge e^{-t\lambda_k}$.

Next, we are interested in proving that B_u are positive for all u. Note that

$$a_{u,1} = \frac{\lambda_k+\lambda_\ell}{2}-\lambda_j, \quad a_{u,2} = \frac{\lambda_j+\lambda_\ell}{2}-\lambda_k, \quad a_{u,3} = \frac{\lambda_j+\lambda_k}{2}-\lambda_\ell,$$

and

$$e^{-t\left(\frac{\lambda_k+\lambda_\ell}{2}+\lambda_j\right)}, \quad e^{-t\left(\frac{\lambda_j+\lambda_\ell}{2}+\lambda_k\right)}, \quad e^{-t\left(\frac{\lambda_j+\lambda_k}{2}+\lambda_\ell\right)},$$

are decreasing.

Now, if $u=1$ or 2, using Lemma 2.5.2 in the Appendix 2.5, we find that

$$\Delta(\beta^{(3)}_{(3,4)},4)\Delta(\beta^{(4)}_u,4) \ge \Delta(\beta^{(3)}_{(u,4)},4)\Delta(\beta^{(4)}_3,4) \ge \Delta(\beta^{(3)}_{(u,3)},4)\Delta(\beta^{(4)}_4,4). \tag{2.2.13}$$

From this, we conclude that

$$\begin{aligned} b_{u,1} &= \Delta(\beta^{(3)}_{(3,4)},4)\Delta(\beta^{(4)}_u,4)\, e^{-t\left(\frac{\lambda_3+\lambda_4}{2}+\lambda_u\right)}, \\ b_{u,2} &= \Delta(\beta^{(3)}_{(u,4)},4)\Delta(\beta^{(4)}_3,4)\, e^{-t\left(\frac{\lambda_u+\lambda_4}{2}+\lambda_3\right)}, \\ b_{u,3} &= \Delta(\beta^{(3)}_{(u,3)},4)\Delta(\beta^{(4)}_4,4)\, e^{-t\left(\frac{\lambda_u+\lambda_3}{2}+\lambda_4\right)}, \end{aligned}$$

are decreasing in $h = 1,2,3$. Note that B_u can be written as $\sum_{h=1}^{3} a_{u,h} b_{u,h}$. Finally, by Lemma 2.1.2,

$$B_u = \sum_{h=1}^{3} a_{u,h}\, b_{u,h} \geq \frac{1}{3}\left(\sum_{h=1}^{3} a_{u,h}\right)\left(\sum_{h=1}^{3} b_{u,h}\right) = 0,$$

since $\sum_{h=1}^{3} a_{u,h} = 0$. Now, if $u = 3$ or 4, we have that

$$\Delta(\beta^{(3)}_{(2,u)},4)\Delta(\beta^{(4)}_{1},4) \geq \Delta(\beta^{(3)}_{(1,2)},4)\Delta(\beta^{(4)}_{u},4), \tag{2.2.14}$$

$$\Delta(\beta^{(3)}_{(1,u)},4)\Delta(\beta^{(4)}_{2},4) \geq \Delta(\beta^{(3)}_{(1,2)},4)\Delta(\beta^{(4)}_{u},4), \tag{2.2.15}$$

and if $\beta^{(3)}_{(1,u)} - \beta^{(4)}_{2} < 0$,

$$\Delta(\beta^{(3)}_{(2,u)},4)\Delta(\beta^{(4)}_{1},4) \geq \Delta(\beta^{(3)}_{(1,u)},4)\Delta(\beta^{(4)}_{2},4). \tag{2.2.16}$$

The proofs of (2.2.14)-(2.2.16) are given in Lemma 2.5.3 in the Appendix 2.5.

It is easy to check that $a_{u,3} = -(a_{u,1} + a_{u,2}) < 0$ and if $\beta^{(3)}_{1,u} - \beta^{(4)}_{2} > 0$ then

$$\begin{aligned} B_u = \sum_{h=1}^{3} a_{u,h} b_{u,h} &\geq min\{b_{u,1}, b_{u,2}\}(a_{u,1} + a_{u,2}) + a_{u,3}\, b_{u,3} \\ &= -a_{u,3}\Big(min\{b_{u,1}, b_{u,2}\} - b_{u,3}\Big) \geq 0, \end{aligned}$$

where

$$\begin{aligned} b_{u,1} &= \Delta(\beta^{(3)}_{(2,u)},4)\Delta(\beta^{(4)}_{1},4)\, e^{-t\left(\frac{\lambda_2+\lambda_u}{2}+\lambda_1\right)}, \\ b_{u,2} &= \Delta(\beta^{(3)}_{(1,u)},4)\Delta(\beta^{(4)}_{2},4)\, e^{-t\left(\frac{\lambda_1+\lambda_u}{2}+\lambda_2\right)}, \\ b_{u,3} &= \Delta(\beta^{(3)}_{(1,2)},4)\Delta(\beta^{(4)}_{u},4)\, e^{-t\left(\frac{\lambda_1+\lambda_2}{2}+\lambda_u\right)}, \end{aligned}$$

and $min\{b_{u,1}, b_{u,2}\} \geq b_{u,3}$ by (2.2.14) and (2.2.15). However, if $\beta^{(3)}_{1,u} - \beta^{(4)}_{2} < 0$, $b_{u,1} \geq b_{u,2} \geq b_{u,3}$ and again by Lemma 2.1.2 $B_u \geq 0$. Hence (2.2.11) holds, which implies that $D^{*}_{3:4} \leq_{hr} D^{*}_{4:4}$ and the proof is complete. □

2.2.2. Simple spacings

We turn to consider the simple spacings of the order statistics where now,

$$\beta^{(i)}_{m_j} = \sum_{\ell=i}^{n} \lambda_{r_\ell}.$$

From (2.1.1) one sees immediately that the p.d.f. of $D_{i:n}$ for $1 \leq i \leq n$ is

$$f_i(t) = \sum_{\mathbf{r}_n} \frac{\prod_{k=1}^{n} \lambda_k}{\prod_{k=1}^{n} \left(\sum_{\ell=k}^{n} \lambda_{r_\ell} \right)} \left(\sum_{\ell=i}^{n} \lambda_{r_\ell} \right) e^{-t \sum_{\ell=i}^{n} \lambda_{r_\ell}},$$

which can be written again as in (2.1.3). Note that the probability $\Delta(\beta_{m_j}^{(i)}, n)$, defined in (2.1.4), is the same in the p.d.f. of $D_{i:n}^*$ and $D_{i:n}$. Therefore, we can apply Lemmas 2.1.1, 2.5.2 and 2.5.3 in order to investigate the hazard rate order of successive simple spacings. This condition is essential to the proof of the next result.

Conjecture 2.2.9. *The successive simple spacings are ordered in the hazard rate ordering, that is,*

$$D_{i:n} \leq_{hr} D_{i+1:n},$$

for $i = 1, \ldots, n-1$.

We give below some advances on this. Specifically, we prove that the second simple spacing is smaller than the third simple spacing according to the hazard rate ordering.

Theorem 2.2.10. *Let* $X_1, \ldots, X_n$ *be independent exponential random variables with rates* $\lambda_1, \ldots, \lambda_n$, *respectively. Then,*

$$D_{2:n} \leq_{hr} D_{3:n} \quad \textit{for any } n.$$

PROOF. We have to show that (2.2.7) holds, where $\beta_{m_j}^{(i)} = \sum_{\ell=i}^{n} \lambda_{r_\ell}$. It is easy to see that for each possible negative element of the matrix $\beta_{m_k}^{(2)} - \beta_{m_j}^{(3)}$, there exists another positive element of form

$$a_{u,1} = \beta_k^{(2)} - \beta_{(j,\ell)}^{(3)} = (\lambda_j + \lambda_\ell) - \lambda_k \geq (\lambda_j + \lambda_k) - \lambda_\ell = \beta_\ell^{(2)} - \beta_{(j,k)}^{(3)} = a_{u,2},$$

since $\beta_k^{(2)} = \sum_{\substack{h=1 \\ h \neq k}}^{n} \lambda_h$, $\beta_{(j,\ell)}^{(3)} = \sum_{\substack{h=1 \\ h \notin \{j,\ell\}}}^{n} \lambda_h$ and $j < k < \ell$. Notice

$$\beta_k^{(2)} + \beta_{(j,\ell)}^{(3)} = (\lambda_j + \lambda_k + \lambda_\ell) + 2 \sum_{\substack{h=1 \\ h \notin \{j,k,\ell\}}}^{n} \lambda_h = \beta_\ell^{(2)} + \beta_{(j,k)}^{(3)},$$

then $e^{-t\left(\beta_k^{(2)} + \beta_{(j,\ell)}^{(3)}\right)} = e^{-t\left(\beta_\ell^{(2)} + \beta_{(j,k)}^{(3)}\right)}$. From equation (2.2.10) and by Lemma 2.1.2, we know

$$\sum_{h=1}^{2} a_{u,h} b_{u,h} \geq \frac{1}{2} \left(\sum_{h=1}^{2} a_{u,h} \right) \left(\sum_{h=1}^{2} b_{u,h} \right) = \lambda_j \left(\sum_{h=1}^{2} b_{u,h} \right) \geq 0,$$

where $b_{u,1} = \Delta(\beta_k^{(2)}, n)\Delta(\beta_{(j,\ell)}^{(3)}, n)$ and $b_{u,2} = \Delta(\beta_\ell^{(2)}, n)\Delta(\beta_{(j,k)}^{(3)}, n)$. This proves the required result. □

Next, we show that the successive simple spacings are increasing in hazard rate ordering for $n = 4$. Hu et al.[31] proved that $D_{1:n} \leq_{lr} D_{2:n}$, and by Theorem 2.2.10 we know that $D_{2:n} \leq_{hr} D_{3:n}$, so we have to show $D_{3:4} \leq_{hr} D_{4:4}$.

Theorem 2.2.11. *Under the same assumptions as those in Theorem 2.2.10, then*

$$D_{3:4} \leq_{hr} D_{4:4}.$$

PROOF. We have to show that (2.2.11) holds where $\beta_{m_j}^{(i)} = \sum_{\ell=i}^{n} \lambda_{r_\ell}$. Here, the matrix of $\beta_{m_k}^{(3)} - \beta_{m_j}^{(4)}$ is

$$\begin{pmatrix} \lambda_3+\lambda_4-\lambda_1 & \lambda_3+\lambda_4-\lambda_2 & \lambda_4 & \lambda_3 \\ \lambda_2+\lambda_4-\lambda_1 & \lambda_4 & \lambda_2+\lambda_4-\lambda_3 & \lambda_2 \\ \lambda_2+\lambda_3-\lambda_1 & \lambda_3 & \lambda_2 & \lambda_2+\lambda_3-\lambda_4 \\ \lambda_4 & \lambda_1+\lambda_4-\lambda_2 & \lambda_1+\lambda_4-\lambda_3 & \lambda_1 \\ \lambda_3 & \lambda_1+\lambda_3-\lambda_2 & \lambda_1 & \lambda_1+\lambda_3-\lambda_4 \\ \lambda_2 & \lambda_1 & \lambda_1+\lambda_2-\lambda_3 & \lambda_1+\lambda_2-\lambda_4 \end{pmatrix} \tag{2.2.17}$$

and we can use the same approach as in the proof of Theorem 2.2.10. It is easy to check that there are only four possible negative coefficients $a_{u,2} = \lambda_j + \lambda_k - \lambda_\ell$ for $j < k < \ell$ and $u \notin \{j,k,\ell\}$. Now, we consider the term $a_{u,1} = \lambda_j + \lambda_\ell - \lambda_k \geq 0$ for $u = 1,\ldots,4$. Notice that $\exp\left\{-t\left(\beta_{(j,\ell)}^{(3)} + \beta_k^{(4)}\right)\right\} = \exp\left\{-t\left(\beta_{(j,k)}^{(3)} + \beta_\ell^{(4)}\right)\right\}$. From equation (2.2.13)) and (2.2.15) we have

$$b_{u,1} = \Delta(\beta_{(j,\ell)}^{(3)},4)\Delta(\beta_k^{(4)},4) \geq \Delta(\beta_{(j,k)}^{(3)},4)\Delta(\beta_\ell^{(4)},4) = b_{u,2}.$$

Hence, by Lemma 2.1.2

$$\sum_{h=1}^{2} a_{u,h}b_{u,h} \geq \frac{1}{2}\left(\sum_{h=1}^{2} a_{u,h}\right)\left(\sum_{h=1}^{2} b_{u,h}\right) = \lambda_j\left(\sum_{h=1}^{2} b_{u,h}\right) \geq 0.$$

This proves the required result. □

The results of this part are mainly based on Torrado et al. [97].

2.3. THE TWO SAMPLE PROBLEM

Let $X_1,\ldots,X_n$ be a set of independent exponential random variables and $Y_1,\ldots,Y_n$ be another set of independent exponential random variables. As we mentioned at the beginning of this chapter, in software reliability, failure times of a software program are modeled as order statistics of inid exponential random variables. Then, we can

consider that $X_{i:n}$ and $Y_{i:n}$ are the failure times of two software programs, and $D_{i:n}$ and $C_{i:n}$ are the i'th times elapsed between software failures, respectively. A natural question is to examine whether the first software program is better than the second one in some stochastic sense.

Many researchers have considered the problem of comparing the spacings of non-identical independent exponential random variables with those corresponding to independent and identically distributed exponential random variables according to different stochastic orderings. Because of the complicated nature of the problem, not much work has been done when the two samples are nonidentical independent exponential random variables. Pledger and Proschan [78] showed that the i'th normalized spacing of a sample of size n from heterogeneous exponential population is stochastically larger than the i'th normalized spacing of a sample of size n whose distribution is the average of the distributions in the heterogeneous case.

Theorem 2.3.1 (Pledger and Proschan [78]). *Let $X_1, X_2, \ldots, X_n$ be independent exponential random variables with X_i having hazard rate λ_i $i = 1, \ldots, n$. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from an exponential distribution with common hazard rate $n\overline{\lambda} = \sum_{i=1}^{n} \lambda_i$. Then*

$$C^*_{i:n} \leq_{st} D^*_{i:n},$$

for $i = 1, \ldots, n$, where $C^*_{i:n} = (n-i+1)(Y_{i:n} - Y_{i-1:n})$ and $D^*_{i:n} = (n-i+1)(X_{i:n} - X_{i-1:n})$ are the i'th normalized spacing from Y_i's and X_i's, respectively, with $Y_{0:n} \equiv 0$ and $X_{0:n} \equiv 0$.

The above result give a nice bound for the survival function of normalized spacings from independent, heterogeneous exponential distributions in terms of the case when they are iid. K& K extended this result from stochastic ordering to likelihood ratio ordering.

Theorem 2.3.2 (K&K [43]). *Under the same assumptions as those in Theorem 2.3.1, then*

$$C^*_{i:n} \leq_{lr} D^*_{i:n},$$

for $i = 1, \ldots, n$.

Kochar and Rojo [44] further strengthened Theorem 2.3.2 to multivariate likelihood ratio order.

Theorem 2.3.3 (Kochar and Rojo [44]). *Under the same assumptions as those in Theorem 2.3.1, then*

$$(C^*_{1:n}, \ldots, C^*_{n:n}) \leq_{lr} (D^*_{1:n}, \ldots, D^*_{n:n}).$$

Kochar and Xu [45] provided necessary and sufficient conditions for stochastically comparing according to likelihood ratio ordering when $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n from an exponential distribution with common hazard rate λ which can differ from $\overline{\lambda}$.

Theorem 2.3.4 (Kochar and Xu [45])**.** *Let $X_1, X_2, \ldots, X_n$ be independent exponential random variables with X_i having hazard rate λ_i $i = 1, \ldots, n$. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from an exponential distribution with common hazard rate λ. Then, for $i \geq 2$,*

$$C_{i:n} \leq_{lr} D_{i:n} \Leftrightarrow (n-i+1)\,\lambda \geq \frac{\sum\limits_{j \in \boldsymbol{r}_n} p_j \left(\sum\limits_{j=i}^{n} \lambda(r_j) \right)^2}{\sum\limits_{j \in \boldsymbol{r}_n} p_j \left(\sum\limits_{j=i}^{n} \lambda(r_j) \right)},$$

for $i = 1, \ldots, n$, where

$$p_j = \frac{\prod\limits_{k=1}^{n} \lambda_k}{\prod\limits_{k=1}^{n} \left(\sum\limits_{j=k}^{n} \lambda(r_j) \right)}.$$

Some of these researchers have investigated the effect on the survival function, the hazard rate function and other characteristics of the time to failure of the spacings when we switch the vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ to another vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$. Pledger and Proschan [78] proved that, in general, the survival function of $D^*_{i:n}$ is not Schur-convex in $(\lambda_1, \ldots, \lambda_n)$. Note that, from Definition 1.4.9, this means that in general, if $\boldsymbol{\theta} \leq^m \boldsymbol{\lambda}$ then $C^*_{i:n} \not\leq_{st} D^*_{i:n}$. However, K&K [43] proved that the survival function of $D^*_{2:n}$ is Schur-convex in $(\lambda_1, \ldots, \lambda_n)$ and, in general, its hazard rate is not Schur-concave, although for $n = 2$, the hazard rate of the second normalized spacing is Schur-concave, i.e., if $\boldsymbol{\theta} \leq^m \boldsymbol{\lambda}$ then $C^*_{2:2} \leq_{hr} D^*_{2:2}$.

Next, we study conditions which are different to that of majorization, which was defined in Subsection 1.4.1, under which normalized and simple spacings are ordered in the likelihood ratio ordering. First, we need an important result and a lemma which is a consequence of Lemma 2.1.1.

Theorem 2.3.5. *Let $X_1, \ldots, X_n$ be independent exponential random variables such that X_i has hazard rate λ_i for $i = 1, \ldots, n$, and $Y_1, \ldots, Y_n$ be independent exponential random variables such that Y_i has hazard rate θ_i for $i = 1, \ldots, n$. Then,*

$$C_{i:n} \leq_{lr} D_{i:n} \Leftrightarrow C^*_{i:n} \leq_{lr} D^*_{i:n},$$

for $i = 1, \ldots, n$.

PROOF. It is easy to see that $D^*_{i:n} = \varphi_i(D_{i:n})$ where $\varphi_i(x) = (n-i+1)x$ is an increasing function. If $C_{i:n} \leq_{lr} D_{i:n}$, then from Theorem 1.C.8. in [88] we get that $C^*_{i:n} \leq_{lr} D^*_{i:n}$, and viceversa, since $\varphi^{-1}(x)$ is also an increasing function. □

Lemma 2.3.6. *Let $\Delta(\beta^{(i)}_{m_j}, n)$ be as defined in (2.1.4). Suppose that m_1 and m_2 are two subsets of $\{1,\ldots,n\}$ of size $n-i+1$ $(1 < i \leq n)$ and having all but one element in common. Denote the different element in m_1 by a_1 and that in m_2 by a_2. Then*

$$\beta^{(i)}_{m_1}\Delta(\beta^{(i)}_{m_1}, n) \geq \beta^{(i)}_{m_2}\Delta(\beta^{(i)}_{m_2}, n) \qquad \text{if} \quad \lambda_{a_2} \geq \lambda_{a_1}.$$

PROOF. Let $c_1,\ldots,c_{i-1}$ be the common elements, then from Lemma 2.1.1, we have

$$\beta^{(i)}_{m_1}\Delta(\beta^{(i)}_{m_1}, n) = \left(\lambda(a_1) + \sum_{j=1}^{i-1} c_i\right)\Delta(\beta^{(i)}_{m_1}, n) \geq \left(\lambda(a_2) + \sum_{j=1}^{i-1} c_i\right)\Delta(\beta^{(i)}_{m_2}, n) = \beta^{(i)}_{m_2}\Delta(\beta^{(i)}_{m_2}, n).$$

□

Now we can establish likelihood ratio ordering between simple spacings from two heterogeneous exponential samples.

Theorem 2.3.7. *Let $X_1,\ldots,X_n$ be independent exponential random variables such that X_i has hazard rate λ_i for $i = 1,\ldots,n$, and $Y_1,\ldots,Y_n$ be independent exponential random variables such that Y_i has hazard rate θ_i for $i = 1,\ldots,n$. If $\alpha^{(i)}_{\min} = \min_{1\leq m_j\leq M_i} \alpha^{(i)}_{m_j} \geq (n-i+1)\overline{\lambda}$, where $\alpha^{(i)}_{m_j} = \sum_{\ell=i}^{n} \theta_{r_\ell}$ and $n\overline{\lambda} = \sum_{i=1}^{n} \lambda_i$. Then*

$$C_{i:n} \leq_{lr} D_{i:n},$$

for $i = 1,\ldots,n$, where $D_{i:n}$ and $C_{i:n}$ are the i'th simple spacing from X_i's and Y_i's, respectively.

PROOF. Observing equation (2.1.3), note that $C_{i:n} \leq_{lr} D_{i:n}$ if and only if

$$\frac{f_{D_{i:n}}(t)}{f_{C_{i:n}}(t)} = \frac{\sum_{k=1}^{M_i} \Delta(\beta^{(i)}_{m_k}, n)\,\beta^{(i)}_{m_k}\, e^{-t\beta^{(i)}_{m_k}}}{\sum_{j=1}^{M_i} \Delta(\alpha^{(i)}_{m_j}, n)\,\alpha^{(i)}_{m_j}\, e^{-t\alpha^{(i)}_{m_j}}},$$

is increasing in t, where $\beta^{(i)}_{m_j} = \sum_{\ell=i}^{n} \lambda_{r_\ell}$ and $\alpha^{(i)}_{m_j} = \sum_{\ell=i}^{n} \theta_{r_\ell}$. Differentiating this equation with respect to t, we have to prove

$$\sum_{k=1}^{M_i}\sum_{j=1}^{M_i} \Delta(\beta^{(i)}_{m_k}, n)\Delta(\alpha^{(i)}_{m_j}, n)\,\beta^{(i)}_{m_k}\alpha^{(i)}_{m_j}\, e^{-t\left(\beta^{(i)}_{m_k}+\alpha^{(i)}_{m_j}\right)}\left(\alpha^{(i)}_{m_j} - \beta^{(i)}_{m_k}\right) \geq 0. \tag{2.3.18}$$

We suppose without loss of generality that the $\beta_{m_k}^{(i)}$'s are in increasing order. By Lemma 2.3.6, we know that $\beta_{m_k}^{(i)}\Delta(\beta_{m_k}^{(i)},n)$'s are in decreasing order, and it is easy to see that $e^{-t\beta_{m_k}^{(i)}}$ and $\left(\alpha_{m_j}^{(i)}-\beta_{m_k}^{(i)}\right)$ are in decreasing order also. Then, by Lemma 2.1.2, we have

$$\sum_{k=1}^{M_i}\sum_{j=1}^{M_i}\Delta(\beta_{m_k}^{(i)},n)\Delta(\alpha_{m_j}^{(i)},n)\,\beta_{m_k}^{(i)}\alpha_{m_j}^{(i)}\,e^{-t\left(\beta_{m_k}^{(i)}+\alpha_{m_j}^{(i)}\right)}\left(\alpha_{m_j}^{(i)}-\beta_{m_k}^{(i)}\right)\geq$$
$$\left(\sum_{k=1}^{M_i}\beta_{m_k}^{(i)}\Delta(\beta_{m_k}^{(i)},n)\,e^{-t\beta_{m_k}^{(i)}}\right)\sum_{j=1}^{M_i}\alpha_{m_j}^{(i)}\Delta(\alpha_{m_j}^{(i)},n)\,e^{-t\alpha_{m_j}^{(i)}}\sum_{k=1}^{M_i}\left(\alpha_{m_j}^{(i)}-\beta_{m_k}^{(i)}\right),$$

where

$$\sum_{k=1}^{M_i}\left(\alpha_{m_j}^{(i)}-\beta_{m_k}^{(i)}\right)=M_i\alpha_{m_j}^{(i)}-\sum_{k=1}^{M_i}\beta_{m_k}^{(i)}=\binom{n}{n-i+1}\alpha_{m_j}^{(i)}-\binom{n-1}{n-i}\sum_{i=1}^{n}\lambda_i\geq 0,$$

if and only if

$$\alpha_{m_j}^{(i)}\geq\binom{n-1}{n-i}\binom{n}{n-i+1}^{-1}\sum_{i=1}^{n}\lambda_i=(n-i+1)\overline{\lambda}.$$

Hence, the required result follows since $\alpha_{\min}^{(i)}\geq(n-i+1)\overline{\lambda}$ for $i=1,\ldots,n$. □

A natural question is to examine if the condition of Theorem 2.3.7 implies majorization and viceversa. The following examples show that, in general, this is not true.

Example 2.3.8 If $\boldsymbol{\theta}=(40,10,1)$ and $\boldsymbol{\lambda}=(40,5.5,5.5)$, it is easy to check that $\boldsymbol{\theta}\geq^m\boldsymbol{\lambda}$, however, for $i=2$,

$$\alpha_{\min}^{(2)}=\min_{1\leq m_j\leq M_2}\alpha_{m_j}^{(2)}=11<34=(n-i+1)\overline{\lambda}=(n-i+1)\overline{\theta}.$$

Note that, in this case, the normalized spacings are not ordered in the hazard rate ordering (see again K&K[43]). ◁

Example 2.3.9 If $\boldsymbol{\theta}=(40,10,1)$ and $\boldsymbol{\lambda}=(5.5,5.5,4)$, for $i=2$ we get

$$\alpha_{\min}^{(2)}=11>10=(n-i+1)\overline{\lambda}\,,$$

and $\boldsymbol{\theta}\not\geq^m\boldsymbol{\lambda}$. ◁

Remark 2.3.10 Let $\{\theta_{(1)},\ldots,\theta_{(n)}\}$ denote the increasing arrangement of θ_i for $i=1,\ldots,n$. It is easy to check that

$$\alpha_{\min}^{(i)}=\sum_{j=1}^{n-i+1}\theta_{(j)}\,,\tag{2.3.19}$$

and

$$\overline{\theta} \geq \frac{\alpha_{\min}^{(2)}}{n-1} \geq \ldots \geq \frac{\alpha_{\min}^{(i)}}{n-i+1} \geq \ldots \geq \frac{\alpha_{\min}^{(n-1)}}{2} \geq \theta_{(1)}. \tag{2.3.20}$$

For $i = n$, from Theorem 2.3.7, we know that if $\theta_{(1)} \geq \overline{\lambda}$ then $C_{n:n} \leq_{lr} D_{n:n}$. Now, by equation (2.3.20) we get that if $\theta_{(1)} \geq \overline{\lambda}$ then $C_{i:n} \leq_{lr} D_{i:n}$ for $i = 1, \ldots, n$. Even more, if we fix i, the condition $\alpha_{\min}^{(i)} \geq (n-i+1)\overline{\lambda}$ of Theorem 2.3.7 implies not only $C_{i:n} \leq_{lr} D_{i:n}$ but also $C_{j:n} \leq_{lr} D_{j:n}$ for $j = 1, \ldots, i$.

Note that for $i = 1$, $X_{1:n} = D_{1:n} = D_{1:n}^*$, and from Theorem 2.3.7 we have

$$\sum_{i=1}^{n} \theta_i \geq \sum_{i=1}^{n} \lambda_i \Rightarrow Y_{1:n} \leq_{lr} X_{1:n},$$

which it is well known since $X_{1:n} \sim \exp(\lambda_1 + \cdots + \lambda_n)$ and $Y_{1:n} \sim \exp(\theta_1 + \cdots + \theta_n)$. ◁

Corollary 2.3.11. *Let $X_1, \ldots, X_n$ be independent exponential random variables such that X_i has hazard rate λ_i for $i = 1, \ldots, n$, and $Y_1, \ldots, Y_n$ be a random sample of size n from an exponential distribution with common hazard rate θ. If $\theta \geq \overline{\lambda}$, then*

$$C_{i:n} \leq_{lr} D_{i:n},$$

for $i = 1, \ldots, n$.

PROOF. It is easy to see that $\alpha_{m_j}^{(i)} = (n-i+1)\theta$ for all m_j, since $Y_1, \ldots, Y_n$ have the same hazard rate. Then, (2.3.18) holds since $\theta \geq \overline{\lambda} \Leftrightarrow \alpha_{m_j}^{(i)} = (n-i+1)\theta \geq (n-i+1)\overline{\lambda}$. □

Note that Theorem 2.3.2 ok K&K can be seen as a particular case of the above Corollary, when $\theta = \overline{\lambda}$. In order to illustrate the performance of the above result, we present here some interesting special cases. Let $X_1, \ldots, X_n$ be independent exponential random variables such that X_i has hazard rate λ_i for $i = 1, \ldots, n$, and $Y_1, \ldots, Y_n$ be a random sample of size n from an exponential distribution with common hazard rate θ. Suppose that $\lambda_1 = \ldots = \lambda_n = \lambda$, from Corollary 2.3.11, it follows that $C_{i:n} \leq_{lr} D_{i:n}$, if $\theta \geq \lambda$, which is a well known result in the literature. Another interesting special case is the following.

Proposition 2.3.12. *Let $X_1, \ldots, X_n$ be independent exponential random variables such that X_i has hazard rate λ_i for $i = 1, \ldots, n$, $Y_1, \ldots, Y_n$ be a random sample of size n from an exponential distribution with common hazard rate $\lambda_{(n)} = \max\{\lambda_1, \ldots, \lambda_n\}$, and $Z_1, \ldots, Z_n$ be a random sample of size n from an exponential distribution with common hazard rate $\lambda_{(1)} = \min\{\lambda_1, \ldots, \lambda_n\}$. Then*

$$C_{i:n} \leq_{lr} D_{i:n} \leq_{lr} H_{i:n},$$

for $i = 1, \ldots, n$ where $C_{i:n}$, $D_{i:n}$, $H_{i:n}$ denote the i'th simple spacings of Y_i's, X_i's and Z_i's, respectively.

PROOF. It is easy to check that $\lambda_{(n)} \geq \overline{\lambda}$. Due to Corollary 2.3.11, it follows that $C_{i:n} \leq_{lr} D_{i:n}$ for $i = 1,\ldots,n$. By (2.3.20), we know that $\beta_{min}^{(i)} \geq (n-i+1)\lambda_{(1)}$ for all i, then from Theorem 2.3.7 we get $D_{i:n} \leq_{lr} H_{i:n}$ for $i = 1,\ldots,n$. □

This result is of interest because it provides upper and lower bounds for the survival and the hazard rate functions since the likelihood ratio order implies the usual stochastic and the hazard rate orders.

Example 2.3.13 Assume that $(\lambda_{(1)},\lambda_{(2)},\lambda_{(3)}) = (0.9,1.0,4.0)$. Figure 2.3 gives an illustration of the above result, where one can see the survival function of the second simple spacing from a heterogeneous exponential random sample with hazard rate $(\lambda_{(1)},\lambda_{(2)},\lambda_{(3)}) = (0.9,1.0,4.0)$. This survival function is bounded by the survival function of the second simple spacing from a exponential random sample with hazard rate $\lambda_{(1)} = 0.9$ and by the survival function of the second simple spacing from a exponential random sample with hazard rate $\lambda_{(3)} = 4$. Even more, we can consider as the lower bound the survival function of the second simple spacing from a exponential random sample with hazard rate $\overline{\lambda} = 1.967$. ◁

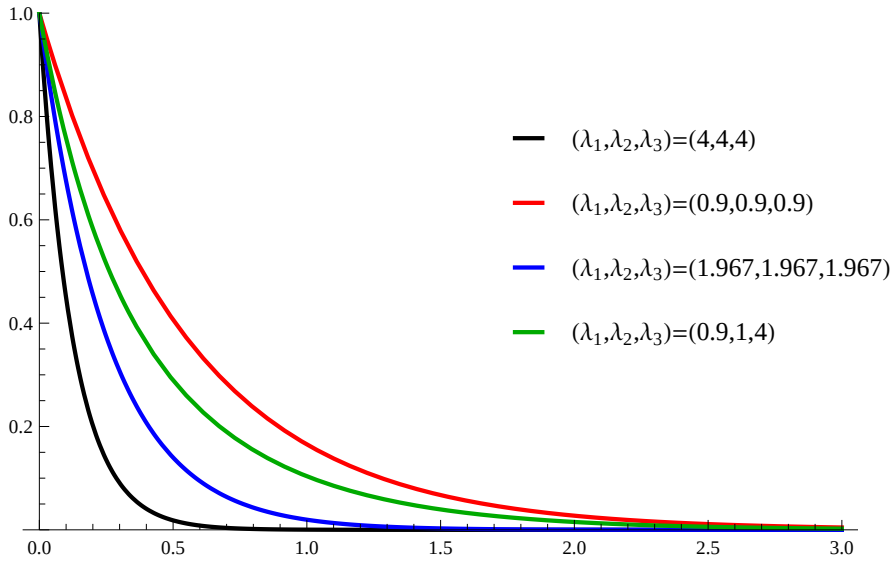


Figure 2.3: The survival curves with different parameters.

Bagai and Kochar [2] proved that if $X \leq_{hr} Y$ and either F or G is DFR (decreasing failure rate), then $X \leq_{\text{disp}} Y$. It is known that spacings of independent heterogeneous exponential random variables have DFR distributions (cf. K&K [43]) and that the likelihood ratio order implies the hazard rate order. Combining these observations, we have proved the following corollary.

Corollary 2.3.14. *Under the same assumptions as those in Theorem 2.3.7,*

$$C_{i:n} \leq_{\text{disp}} D_{i:n},$$

for $i = 1, \ldots, n$.

Consequences of Corollary 2.3.14 are that $\text{Var}(C_{i:n}) \leq \text{Var}(D_{i:n})$ for $i = 1, \ldots, n$.

2.3.1. Applications to a multiple-outlier model

By considering the multiple-outlier model as a special case in the independent and non identically distributed framework, we present results on simple spacings from a multiple-outlier exponential model. Note that from Theorem 2.3.5, we can apply these results to normalized spacings from a multiple-outlier exponential model.

In this section, we consider the special case when $X_1, \ldots, X_n$ are independent exponential random variables such that X_i has hazard rate λ for $i = 1, \ldots, p$ and X_j has hazard rate λ_* for $j = p+1, \ldots, n$, where two samples are independent. The simple spacings and normalized spacings from a multiple-outlier exponential model are, respectively, defined by

$$D_{i:n}(p,q;\lambda,\lambda_*) = X_{i:n} - X_{i-1:n} \quad \text{and} \quad D^*_{i:n}(p,q;\lambda,\lambda_*) = (n-i+1)\, D_{i:n}(p,q;\lambda,\lambda_*),$$

for $i = 1, \ldots, n$, with $X_{0:n} \equiv 0$, $q = n - p \geq 1$ and $p \geq 1$. To simplify notation, we shall write $D_{i:n}(p,q)$ and $D_{i:n}(\lambda,\lambda_*)$ instead of $D_{i:n}(p,q;\lambda,\lambda_*)$ when there is no ambiguity, and the dependence of spacings on the parameters (p,q) and (λ,λ_*) are emphasized, respectively. The notation $D_{i:n}(p,q)$ and $D_{i:n}(\lambda,\lambda_*)$ have a similar interpretation.

As we mentioned in Theorem 2.2.3 in Section 2.2, Khaledi and Kochar [37] proved that

$$D^*_{i:n}(n-1,1) \leq_{hr} D^*_{i+1:n}(n-1,1), \quad \text{for } i = 1, \ldots, n-1,$$

in a single-outlier exponential model. Wen et al. [99] established the likelihood ratio ordering of simple spacings from a multiple-outlier exponential model (see Theorem 2.2.4), that is,

$$D_{i:n}(p,q) \leq_{lr} D_{i+1:n}(p,q), \quad \text{for } p \geq 1, q \geq 1 \text{ and } i = 1, \ldots, n-1.$$

Hu et al. [32] also investigated stochastic comparisons of simple spacings from a multiple-outlier exponential model. They proved, for $\lambda_1 \leq \lambda_* \leq \lambda_2$,

$$\Big(D_{1:n}(\lambda_2,\lambda_*), \ldots, D_{n:n}(\lambda_2,\lambda_*)\Big) \leq_{lr} \Big(D_{1:n}(\lambda_1,\lambda_*), \ldots, D_{n:n}(\lambda_1,\lambda_*)\Big),$$

with $p,q \geq 2$. Since the multivariate likelihood ratio order is closed under marginalization (see Shaked and Shanthikumar[88]), it holds that, for $\lambda_1 \leq \lambda_* \leq \lambda_2$,

$$D_{i:n}(\lambda_2,\lambda_*) \leq_{lr} D_{i:n}(\lambda_1,\lambda_*), \quad \text{for } i=1,\ldots,n. \tag{2.3.21}$$

In the following example, we show that (2.3.21) is a special case of Theorem 2.3.7.

Example 2.3.15 Suppose that $\lambda_1 \leq \lambda_* \leq \lambda_2$. We will show that $\alpha_{min}^{(i)} \geq (n-i+1)\overline{\lambda}$, where

$$\overline{\lambda} = \frac{p\lambda_1+(n-p)\lambda_*}{n}, \quad \text{and}$$
$$\alpha_{min}^{(i)} = \begin{cases} (n-i+1)\lambda_*, & \text{if } i \geq p+1, \\ (n-p)\lambda_*+(p-i+1)\lambda_2, & \text{if } i < p+1. \end{cases} \tag{2.3.22}$$

Then, from Theorem 2.3.7, we get that $D_{i:n}(\lambda_2,\lambda_*) \leq_{lr} D_{i:n}(\lambda_1,\lambda_*)$, for $i=1,\ldots,n$. When $i \geq p+1$,

$$\alpha_{min}^{(i)} \geq (n-i+1)\overline{\lambda} \Leftrightarrow n\lambda_* \geq p\lambda_1+(n-p)\lambda_* \Leftrightarrow \lambda_* \geq \lambda_1.$$

And, when $i < p+1$, $\alpha_{min}^{(i)} \geq (n-i+1)\overline{\lambda} \Leftrightarrow (n-p)\lambda_*+(p-i+1)\lambda_2 \geq (n-i+1)\overline{\lambda}$. As $\lambda_2 \geq \lambda_*$, it is easy to see that, if $i < p+1$,

$$(n-p)\lambda_*+(p-i+1)\lambda_2 = (n-i+1)\lambda_* \geq (n-i+1)\overline{\lambda} \Leftrightarrow \lambda_* \geq \lambda_1.$$

Hence, $D_{i:n}(\lambda_2,\lambda_*) \leq_{lr} D_{i:n}(\lambda_1,\lambda_*)$, for $i=1,\ldots,n$. ◁

Using again Theorem 2.3.7, we give below a similar result to (2.3.21) when the number of exponential random variables with hazard rate λ_1 and λ_* can be changed.

Theorem 2.3.16. *Let $X_1,\ldots,X_n$ follow the multiple-outlier model with parameters λ_1 and λ_* and let $Y_1,\ldots,Y_n$ follow the multiple-outlier model with parameters λ_2 and λ_*. If $\lambda_1 \leq \lambda_* \leq \lambda_2$, then*

i) $D_{i:n}(p,q;\lambda_2,\lambda_*) \leq_{lr} D_{i:n}(p+k_1,q-k_1;\lambda_1,\lambda_*)$, *with* $1 \leq k_1 \leq q$ *and*

ii) $D_{i:n}(p,q;\lambda_2,\lambda_*) \leq_{lr} D_{i:n}(p-k_2,q+k_2;\lambda_1,\lambda_*)$, *with* $1 \leq k_2 \leq p$,

where $q=n-p \geq 1$, $p \geq 1$.

PROOF. We have to show that $\alpha_{min}^{(i)} \geq (n-i+1)\overline{\lambda}$ and then, from Theorem 2.3.7 we will conclude that the result follows. It is easy to see that (2.3.22) holds.

i) In this case, $n\overline{\lambda} = (p+k_1)\lambda_1 + (q-k_1)\lambda_*$, with $1 \le k_1 \le q$. When $i \ge p+1$,

$$\alpha_{min}^{(i)} \ge (n-i+1)\overline{\lambda} \Leftrightarrow (n-i+1)\lambda_* \ge (n-i+1)\overline{\lambda} \Leftrightarrow (\lambda_* - \lambda_1)(p+k_1) \ge 0 \Leftrightarrow \lambda_* \ge \lambda_1.$$

And when $i < p+1$, $\alpha_{min}^{(i)} \ge (n-i+1)\overline{\lambda} \Leftrightarrow (n-p)\lambda_* + (p-i+1)\lambda_2 \ge (n-i+1)\overline{\lambda}$. As $\lambda_2 \ge \lambda_*$, then

$$\alpha_{min}^{(i)} \ge (n-i+1)\overline{\lambda} \Leftrightarrow (n-i+1)\lambda_* \ge (n-i+1)\overline{\lambda} \Leftrightarrow \lambda_* \ge \lambda_1.$$

ii) In this case, $n\overline{\lambda} = (p-k_2)\lambda_1 + (q+k_2)\lambda_*$, where $1 \le k_2 \le p$. As before, it is easy to check that

$$\alpha_{min}^{(i)} \ge (n-i+1)\overline{\lambda} \Leftrightarrow (n-i+1)\lambda_* \ge (n-i+1)\overline{\lambda} \Leftrightarrow (\lambda_* - \lambda_1)(p-k_2) \ge 0 \Leftrightarrow \lambda_* \ge \lambda_1,$$

for $i = 1, \ldots, n$. □

Wen et al. [99] obtained the following result.

Theorem 2.3.17 (Wen et al. [99]). *Let $X_1, \ldots, X_n$ follow the multiple-outlier model with parameters λ and λ_*. If $\lambda \le \lambda_*$, $p \ge 1$ and $q \ge 1$, then*

$$D_{i:n}(p,q) \le_{lr} D_{i:n}(p+1, q-1), \quad \textit{for } i = 1, \ldots, n.$$

We now state the analogue of this last result as a special case of Theorem 2.3.7, when $\lambda \ge \lambda_*$.

Theorem 2.3.18. *Let $X_1, \ldots, X_n$ follow the multiple-outlier model with parameters λ and λ_*. If $\lambda \ge \lambda_*$, $p \ge 1$ and $q \ge 1$, then*

$$D_{i:n}(p-k_2, q+k_2) \ge_{lr} D_{i:n}(p,q) \ge_{lr} D_{i:n}(p+k_1, q-k_1),$$

where $1 \le k_1 \le q$, $1 \le k_2 \le p$ and $i = 1, \ldots, n$.

PROOF. First, we will see that $D_{i:n}(p-k_2, q+k_2) \ge_{lr} D_{i:n}(p,q)$, where $1 \le k_2 \le p$. A trivial verification shows that

$$\overline{\lambda} = \frac{(p-k_2)\lambda + (q+k_2)\lambda_*}{n}, \quad \text{and}$$

$$\alpha_{min}^{(i)} = \begin{cases} (n-i+1)\lambda, & \text{if } i \ge q+1, \\ p\lambda + (n-i+1-p)\lambda_*, & \text{if } i < q+1. \end{cases}$$

It follows immediately that, if $i \ge q+1$,

$$\begin{aligned} \alpha_{min}^{(i)} \ge (n-i+1)\overline{\lambda} &\Leftrightarrow n\lambda \ge (p-k_2)\lambda + (q+k_2)\lambda_* \\ &\Leftrightarrow (\lambda - \lambda_*)(q+k_2) \ge 0 \Leftrightarrow \lambda \ge \lambda_*. \end{aligned}$$

And, if $i < q+1$, then,

$$\alpha_{min}^{(i)} \geq (n-i+1)\overline{\lambda} \Leftrightarrow p\lambda + (n-i+1-p)\lambda_* \geq \frac{n-i+1}{n}\Big((p-k_2)\lambda + (q+k_2)\lambda_*\Big)$$
$$\Leftrightarrow (\lambda-\lambda_*)\Big(nk_2 + (i-1)(p-k_2)\Big) \geq 0 \Leftrightarrow \lambda \geq \lambda_*.$$

To prove that $D_{i:n}(p,q) \geq_{lr} D_{i:n}(p+k_1, q-k_1)$ where $1 \leq k_1 \leq q$, we get

$$\overline{\lambda} = \frac{p\lambda + q\lambda_*}{n}, \quad \text{and}$$
$$\alpha_{min}^{(i)} = \begin{cases} (n-i+1)\lambda, & \text{if } n-i+1 \leq p+k_1, \\ (p+k_1)\lambda + (n-i+1-p-k_1)\lambda_*, & \text{if } n-i+1 > p+k_1. \end{cases}$$

Clearly, when $n-i+1 \leq p+k_1$,

$$\alpha_{min}^{(i)} \geq (n-i+1)\overline{\lambda} \Leftrightarrow n\lambda \geq p\lambda + q\lambda_* \Leftrightarrow q(\lambda-\lambda_*) \geq 0 \Leftrightarrow \lambda \geq \lambda_*.$$

And when $n-i+1 > p+k_1$, we have

$$\alpha_{min}^{(i)} \geq (n-i+1)\overline{\lambda} \Leftrightarrow (p+k_1)\lambda + (n-i+1-p-k_1)\lambda_* \geq \frac{n-i+1}{n}(p\lambda + q\lambda_*)$$
$$\Leftrightarrow (\lambda-\lambda_*)\Big(nk_1 + p(i-1)\Big) \geq 0 \Leftrightarrow \lambda \geq \lambda_*.$$

Hence, we have proved that $\alpha_{min}^{(i)} \geq (n-i+1)\overline{\lambda} \Leftrightarrow \lambda \geq \lambda_*$, and from Theorem 2.3.7 we get the desired result. □

2.4. SUMMARY AND FUTURE WORK

This chapter is devoted to establishing stochastic comparisons of spacings from one and two samples of heterogeneous exponential random variables.

In the first part of this chapter, we have shown, in the case $n = 4$, that both normalized spacings and simple spacings are ordered according to the hazard rate ordering. We have also established hazard rate ordering between the second and the third simple spacings and normalized spacings for any n and for all λ_i's. As we mentioned en Section 2.2, the results of this part are mainly based on Torrado et al. [97].

The second part of this chapter concerns stochastic comparisons between spacings from two samples of exponential random variables with different scale parameters. We have provided sufficient conditions under which the simple and normalized spacings are ordered according to the likelihood ratio ordering. We also have shown applications of these results to the multiple-outlier exponential model.

When we compare spacings from one sample, our method is based on the fact that in each row of the matrix of differences between the betas, there exist a number of elements that sum to zero. We believe that by studying the structure of these matrices, an adequate form of applying Chebyshev's inequality to the general solution of the conjecture of K&K can be found.

To end this chapter, we make the following conjectures. Let $X_1,\ldots,X_n$ be independent exponential random variables with X_i having failure rate λ_i for each i. Let $\Delta(\beta_{m_j}^{(n-1)},n)$ be as defined in (2.1.4). Let $\beta_{m_k}^{(n-1)}=\frac{\lambda_j+\lambda_\ell}{2}$ and $\beta_{m_j}^{(n)}=\lambda_j$ be, if $m_k=(j,\ell)$ and $m_j=j$, respectively. Then, we think that the following inequalities hold,

❶ $\Delta(\beta_{(k,\ell)}^{(n-1)},n)\Delta(\beta_j^{(n)},n)\geq\Delta(\beta_{(j,k)}^{(n-1)},n)\Delta(\beta_\ell^{(n)},n)$,

❷ $\Delta(\beta_{(j,\ell)}^{(n-1)},n)\Delta(\beta_k^{(n)},n)\geq\Delta(\beta_{(j,k)}^{(n-1)},n)\Delta(\beta_\ell^{(n)},n)$,

❸ $\Delta(\beta_{(k,\ell)}^{(n-1)},n)\Delta(\beta_j^{(n)},n)\geq\Delta(\beta_{(j,\ell)}^{(n-1)},n)\Delta(\beta_k^{(n)},n)$ for $j\neq 1$ or $k\neq 2$,

❹ $\Delta(\beta_{(k,\ell)}^{(n-1)},n)\Delta(\beta_j^{(n)},n)\geq\Delta(\beta_{(j,\ell)}^{(n-1)},n)\Delta(\beta_k^{(n)},n)$ if $\beta_{(j,\ell)}^{(n-1)}-\beta_k^{(n)}<0$ for $j=1$ and $k=2$.

Assuming that our conjectures hold, it would then be possible to prove that

$$D^*_{n-1:n}\leq_{hr}D^*_{n:n}\quad\text{and}\quad D_{n-1:n}\leq_{hr}D_{n:n}\qquad\text{for any } n.$$

2.5. APPENDIX: PROOFS

This appendix contains a particular case of Theorem 2.2.7 when $n=4$ and also we show here that equations (2.2.13)-(2.2.16) hold.

Theorem 2.5.1. *Let $X_1,\ldots,X_n$ be independent exponential random variables such that X_i has hazard rate λ_i, for $i=1,\ldots,n$, then $D^*_{2:4}\leq_{hr}D^*_{3:4}$.*

PROOF. We have to show

$$\sum_{j=1}^{M_3}\sum_{k=1}^{M_2}\Delta(\beta_{m_k}^{(2)},4)\Delta(\beta_{m_j}^{(3)},4)\,e^{-t\left(\beta_{m_k}^{(2)}+\beta_{m_j}^{(3)}\right)}\left(\beta_{m_k}^{(2)}-\beta_{m_j}^{(3)}\right)\geq 0. \tag{2.5.23}$$

To do this, we consider the values of $\beta_{m_k}^{(2)}-\beta_{m_j}^{(3)}$ which add zero for each $k=1,\ldots,4$

and $j = 1, \ldots, 6$ in the next matrix transpose.

$$\begin{pmatrix}
\frac{\lambda_2+\lambda_3+\lambda_4}{3} - \frac{\lambda_1+\lambda_2}{2} & \frac{\lambda_1+\lambda_3+\lambda_4}{3} - \frac{\lambda_1+\lambda_2}{2} & \frac{\lambda_1+\lambda_2+\lambda_4}{3} - \frac{\lambda_1+\lambda_2}{2} & \frac{\lambda_1+\lambda_2+\lambda_3}{3} - \frac{\lambda_1+\lambda_2}{2} \\
\frac{\lambda_2+\lambda_3+\lambda_4}{3} - \frac{\lambda_1+\lambda_3}{2} & \frac{\lambda_1+\lambda_3+\lambda_4}{3} - \frac{\lambda_1+\lambda_3}{2} & \frac{\lambda_1+\lambda_2+\lambda_4}{3} - \frac{\lambda_1+\lambda_3}{2} & \frac{\lambda_1+\lambda_2+\lambda_3}{3} - \frac{\lambda_1+\lambda_3}{2} \\
\frac{\lambda_2+\lambda_3+\lambda_4}{3} - \frac{\lambda_1+\lambda_4}{2} & \frac{\lambda_1+\lambda_3+\lambda_4}{3} - \frac{\lambda_1+\lambda_4}{2} & \frac{\lambda_1+\lambda_2+\lambda_4}{3} - \frac{\lambda_1+\lambda_4}{2} & \frac{\lambda_1+\lambda_2+\lambda_3}{3} - \frac{\lambda_1+\lambda_4}{2} \\
\frac{\lambda_2+\lambda_3+\lambda_4}{3} - \frac{\lambda_2+\lambda_3}{2} & \frac{\lambda_1+\lambda_3+\lambda_4}{3} - \frac{\lambda_2+\lambda_3}{2} & \frac{\lambda_1+\lambda_2+\lambda_4}{3} - \frac{\lambda_2+\lambda_3}{2} & \frac{\lambda_1+\lambda_2+\lambda_3}{3} - \frac{\lambda_2+\lambda_3}{2} \\
\frac{\lambda_2+\lambda_3+\lambda_4}{3} - \frac{\lambda_2+\lambda_4}{2} & \frac{\lambda_1+\lambda_3+\lambda_4}{3} - \frac{\lambda_2+\lambda_4}{2} & \frac{\lambda_1+\lambda_2+\lambda_4}{3} - \frac{\lambda_2+\lambda_4}{2} & \frac{\lambda_1+\lambda_2+\lambda_3}{3} - \frac{\lambda_2+\lambda_4}{2} \\
\frac{\lambda_2+\lambda_3+\lambda_4}{3} - \frac{\lambda_3+\lambda_4}{2} & \frac{\lambda_1+\lambda_3+\lambda_4}{3} - \frac{\lambda_3+\lambda_4}{2} & \frac{\lambda_1+\lambda_2+\lambda_4}{3} - \frac{\lambda_3+\lambda_4}{2} & \frac{\lambda_1+\lambda_2+\lambda_3}{3} - \frac{\lambda_3+\lambda_4}{2}
\end{pmatrix} \tag{2.5.24}$$

To simplify the notation let $\beta_u^{(2)} = \frac{\lambda_j+\lambda_k+\lambda_\ell}{3}$ where $u \notin \{j,k,\ell\}$ and $\beta_{(j,k)}^{(3)} = \frac{\lambda_j+\lambda_k}{2}$ be. There is an addition as following in each row in the transpose of the above matrix,

$$\begin{aligned}
&(\beta_u^{(2)} - \beta_{(j,k)}^{(3)}) \quad + \quad (\beta_u^{(2)} - \beta_{(j,\ell)}^{(3)}) \quad + \quad (\beta_u^{(2)} - \beta_{(k,\ell)}^{(3)}) \quad = \\
&\left(\frac{\lambda_j+\lambda_k+\lambda_\ell}{3} - \frac{\lambda_j+\lambda_j}{2}\right) + \left(\frac{\lambda_j+\lambda_k+\lambda_\ell}{3} - \frac{\lambda_j+\lambda_\ell}{2}\right) + \left(\frac{\lambda_j+\lambda_k+\lambda_\ell}{3} - \frac{\lambda_k+\lambda_\ell}{2}\right) = 0,
\end{aligned}$$

for $j,k,\ell = 1,\ldots,4$ and $j < k < \ell$. We are interested in proving

$$\begin{aligned}
&\Delta(\beta_u^{(2)},4)\, e^{-t\left(\frac{\lambda_j+\lambda_k+\lambda_\ell}{3}\right)} \Bigg[\Delta(\beta_{(j,k)}^{(3)},4)\, e^{-t\left(\frac{\lambda_j+\lambda_k}{2}\right)} \left(\beta_u^{(2)} - \beta_{(j,k)}^{3}\right) + \\
&\Delta(\beta_{(j,\ell)}^{(3)},4)\, e^{-t\left(\frac{\lambda_j+\lambda_\ell}{2}\right)} \left(\beta_u^{(2)} - \beta_{(j,\ell)}^{3}\right) + \Delta(\beta_{(k,\ell)}^{(3)},4)\, e^{-t\left(\frac{\lambda_k+\lambda_\ell}{2}\right)} \left(\beta_u^{(2)} - \beta_{(k,\ell)}^{3}\right) \Bigg] \geq 0,
\end{aligned} \tag{2.5.25}$$

for $u = 1,\ldots,4$.

Notice that $a_1 = \beta_u^{(2)} - \beta_{(j,k)}^{(3)}$, $a_2 = \beta_u^{(2)} - \beta_{(j,\ell)}^{(3)}$ and $a_3 = \beta_u^{(2)} - \beta_{(k,\ell)}^{(3)}$ are decreasing in $h = 1,2,3$, and the corresponding exponentials too. It follows from Lemma 2.1.1 that

$$\Delta(\beta_{(j,k)}^{(3)},4) \geq \Delta(\beta_{(j,\ell)}^{(3)},4) \geq \Delta(\beta_{(k,\ell)}^{3},4).$$

Then

$$\begin{aligned}
b_1 &= \Delta(\beta_{(j,k)}^{(3)},4)\, e^{-t\left(\frac{\lambda_j+\lambda_k}{2}\right)}, \\
b_2 &= \Delta(\beta_{(j,\ell)}^{(3)},4)\, e^{-t\left(\frac{\lambda_j+\lambda_\ell}{2}\right)}, \\
b_3 &= \Delta(\beta_{(k,\ell)}^{(3)},4)\, e^{-t\left(\frac{\lambda_k+\lambda_\ell}{2}\right)},
\end{aligned}$$

are decreasing in $h = 1,2,3$. Finally, by Lemma 2.1.2, we conclude that (2.5.25) holds since $\sum_{h=1}^{3} a_i = 0$. We group twelve remaining values of the matrix (2.5.24) in four diagonals.

$$\begin{aligned}
a_{1,2} &= \tfrac{\lambda_1+\lambda_3+\lambda_4}{3} - \tfrac{\lambda_1+\lambda_2}{2} \geq a_{1,3} = \tfrac{\lambda_1+\lambda_2+\lambda_4}{3} - \tfrac{\lambda_1+\lambda_3}{2} \geq a_{1,4} = \tfrac{\lambda_1+\lambda_2+\lambda_3}{3} - \tfrac{\lambda_1+\lambda_4}{2} \\
a_{2,1} &= \tfrac{\lambda_2+\lambda_3+\lambda_4}{3} - \tfrac{\lambda_1+\lambda_2}{2} \geq a_{2,3} = \tfrac{\lambda_1+\lambda_2+\lambda_4}{3} - \tfrac{\lambda_2+\lambda_3}{2} \geq a_{2,4} = \tfrac{\lambda_1+\lambda_2+\lambda_3}{3} - \tfrac{\lambda_2+\lambda_4}{2} \\
a_{3,1} &= \tfrac{\lambda_2+\lambda_3+\lambda_4}{3} - \tfrac{\lambda_1+\lambda_3}{2} \geq a_{3,2} = \tfrac{\lambda_1+\lambda_3+\lambda_4}{3} - \tfrac{\lambda_2+\lambda_3}{2} \geq a_{3,4} = \tfrac{\lambda_1+\lambda_2+\lambda_3}{3} - \tfrac{\lambda_3+\lambda_4}{2} \\
a_{4,1} &= \tfrac{\lambda_2+\lambda_3+\lambda_4}{3} - \tfrac{\lambda_1+\lambda_4}{2} \geq a_{4,2} = \tfrac{\lambda_1+\lambda_3+\lambda_4}{3} - \tfrac{\lambda_2+\lambda_4}{2} \geq a_{4,3} = \tfrac{\lambda_1+\lambda_2+\lambda_4}{3} - \tfrac{\lambda_3+\lambda_4}{2}
\end{aligned} \tag{2.5.26}$$

Firstly we prove that the coefficients $\Delta(\beta^{(2)}_{m_k},4)\Delta(\beta^{(3)}_{m_j},4)$ in (2.5.23) related to the four diagonals are also ordered. We give the proof only for the first diagonal, the other cases are similar. Again by Lemma 2.1.1,

$$\begin{aligned}
\Delta(\beta^{(2)}_2,4)\Delta(\beta^{(3)}_{(1,2)},4) \quad &\geq \quad \Delta(\beta^{(2)}_3,4)\Delta(\beta^{(3)}_{(1,3)},4) \quad \geq \quad \Delta(\beta^{(2)}_4,4)\Delta(\beta^{(3)}_{(1,4)},4) \Leftrightarrow \\
\frac{\lambda_2}{s_4}\Delta(\beta^{(3)}_{(1,2)},4) \quad &\geq \quad \frac{\lambda_3}{s_4}\Delta(\beta^{(3)}_{(1,3)},4) \quad \geq \quad \frac{\lambda_4}{s_4}\Delta(\beta^{(3)}_{(1,4)},4),
\end{aligned}$$

where $s_4 = \sum_{k=1}^{4} \lambda_k$. It is easy to see that exponential are decreasing, so are

$$\begin{aligned}
b_{1,2} &= \Delta(\beta^{(2)}_2,4)\Delta(\beta^{(3)}_{(1,2)},4)\, e^{-t\left(\frac{\lambda_1+\lambda_3+\lambda_4}{3}+\frac{\lambda_1+\lambda_2}{2}\right)}, \\
b_{1,3} &= \Delta(\beta^{(2)}_3,4)\Delta(\beta^{(3)}_{(1,3)},4)\, e^{-t\left(\frac{\lambda_1+\lambda_2+\lambda_4}{3}+\frac{\lambda_1+\lambda_3}{2}\right)}, \\
b_{1,4} &= \Delta(\beta^{(2)}_4,4)\Delta(\beta^{(3)}_{(1,4)},4)\, e^{-t\left(\frac{\lambda_1+\lambda_2+\lambda_3}{3}+\frac{\lambda_1+\lambda_4}{2}\right)}.
\end{aligned}$$

Hence, by Lemma 2.1.2,

$$\sum_{\substack{j=1\\ j\neq 1}}^{4} a_{1,j}\, b_{1,j} \geq \frac{1}{3}\left(\sum_{\substack{j=1\\ j\neq 1}}^{4} a_{1,j}\right)\left(\sum_{\substack{j=1\\ j\neq 1}}^{4} b_{1,j}\right) = \frac{1}{3}\, a_1 b_1.$$

Secondly we will verify that $b_1 \geq b_2 \geq b_3 \geq b_4$. To do this, it sufficient to prove that the corresponding $b_{i,j}$ in the array (2.5.26) are ordered by rows. We give the proof only for the first two rows, so we will see that $b_{1,2} \geq b_{2,1}$, $b_{1,3} \geq b_{2,3}$ and $b_{1,4} \geq b_{2,4}$. It is immediate that

$$
\begin{aligned}
e^{-t\left(\frac{\lambda_1+\lambda_3+\lambda_4}{3}+\frac{\lambda_1+\lambda_2}{2}\right)} &\geq e^{-t\left(\frac{\lambda_2+\lambda_3+\lambda_4}{3}+\frac{\lambda_1+\lambda_2}{2}\right)},\\
e^{-t\left(\frac{\lambda_1+\lambda_2+\lambda_4}{3}+\frac{\lambda_1+\lambda_3}{2}\right)} &\geq e^{-t\left(\frac{\lambda_1+\lambda_2+\lambda_4}{3}+\frac{\lambda_2+\lambda_3}{2}\right)},\\
e^{-t\left(\frac{\lambda_1+\lambda_2+\lambda_3}{3}+\frac{\lambda_1+\lambda_4}{2}\right)} &\geq e^{-t\left(\frac{\lambda_1+\lambda_2+\lambda_3}{3}+\frac{\lambda_2+\lambda_4}{2}\right)}.
\end{aligned}
$$

It follows easily that

$$
\Delta(\beta_2^{(2)},4)\Delta(\beta_{(1,2)}^{(3)},4) \geq \Delta(\beta_1^{(2)},4)\Delta(\beta_{(1,2)}^{(3)},4) \Leftrightarrow \Delta(\beta_2^{(2)},4) = \frac{\lambda_2}{s_4} \geq \frac{\lambda_1}{s_4} = \Delta(\beta_1^{(2)},4),
$$

then

$$
b_{1,2} \geq b_{2,1} = \Delta(\beta_1^{(2)},4)\Delta(\beta_{(1,2)}^{(3)},4)e^{-t\left(\frac{\lambda_2+\lambda_3+\lambda_4}{3}+\frac{\lambda_1+\lambda_2}{2}\right)}.
$$

Now from Lemma 2.1.1

$$
\begin{aligned}
\Delta(\beta_3^{(2)},4)\Delta(\beta_{(1,3)}^{(3)},4) \geq \Delta(\beta_3^{(2)},4)\Delta(\beta_{(2,3)}^{(3)},4) &\Leftrightarrow \Delta(\beta_{(1,3)}^{(3)},4) \geq \Delta(\beta_{(2,3)}^{(3)},4),\\
\Delta(\beta_4^{(2)},4)\Delta(\beta_{(1,4)}^{(3)},4) \geq \Delta(\beta_4^{(2)},4)\Delta(\beta_{(2,4)}^{(3)},4) &\Leftrightarrow \Delta(\beta_{(1,4)}^{(3)},4) \geq \Delta(\beta_{(2,4)}^{(3)},4),
\end{aligned}
$$

then

$$
\begin{aligned}
b_{1,3} \geq b_{2,3} &= \Delta(\beta_3^{(2)},4)\Delta(\beta_{(2,3)}^{(3)},4)\, e^{-t\left(\frac{\lambda_1+\lambda_2+\lambda_4}{3}+\frac{\lambda_2+\lambda_3}{2}\right)},\\
b_{1,4} \geq b_{2,4} &= \Delta(\beta_4^{(2)},4)\Delta(\beta_{(2,4)}^{(3)},4)\, e^{-t\left(\frac{\lambda_1+\lambda_2+\lambda_3}{3}+\frac{\lambda_2+\lambda_4}{2}\right)}.
\end{aligned}
$$

Consequently

$$
b_1 = \sum_{\substack{j=1\\ j\neq 1}}^{4} b_{1,j} \geq \sum_{\substack{j=1\\ j\neq 2}}^{4} b_{2,j} = b_2.
$$

The same reasoning applies to the other cases. Then $b_1 \geq b_2 \geq b_3 \geq b_4$. We define

$$
a_k = \sum_{\substack{j=1\\ j\neq k}}^{4} a_{k,j}\,,
$$

for $k = 1,\ldots,4$. Clearly, $a_1 \geq a_2 \geq a_3 \geq a_4$. Since $\sum_{k=1}^{4} a_k = 0$, by Lemma 2.1.2 and (2.5.25) we conclude that (2.5.23) holds. □

Lemma 2.5.2. *Let $\Delta(\beta_{m_j}^{(i)},n)$ be as in (2.1.4), and $m_j = j$ and $m_k = (j,\ell)$. Then*

$$
\Delta(\beta_{(3,4)}^{(3)},4)\Delta(\beta_u^{(4)},4) \geq \Delta(\beta_{(u,4)}^{(3)},4)\Delta(\beta_3^{(4)},4) \geq \Delta(\beta_{(u,3)}^{(3)},4)\Delta(\beta_4^{(4)},4) \quad \textit{for } u = 1,2.
$$

PROOF. We divide the proof into two parts

(a) $\Delta(\beta^{(3)}_{(3,4)},4)\Delta(\beta^{(4)}_{u},4) \geq \Delta(\beta^{(3)}_{(u,4)},4)\Delta(\beta^{(4)}_{3},4)$,

(b) $\Delta(\beta^{(3)}_{(u,4)},4)\Delta(\beta^{(4)}_{3},4) \geq \Delta(\beta^{(39}_{(u,3)},4)\Delta(\beta^{(4)}_{4},4)$.

We give the proof only for the case $u = 1$, the other case is similar with 1 replaced by 2. After a few manipulations, we have

$$\begin{aligned}
\Delta(\beta^{(3)}_{(3,4)},4)\Delta(\beta^{(4)}_{1},4) &= \left(\frac{\lambda_1\lambda_2^2\lambda_3\lambda_4}{S s_4 s_3 s_2^2 s_1}\right)\frac{s_1+s_2}{(\lambda_1+\lambda_2)(\lambda_1+\lambda_3)(\lambda_1+\lambda_4)}Z_1\,,\\
\Delta(\beta^{(3)}_{(1,4)},4)\Delta(\beta^{(4)}_{3},4) &= \left(\frac{\lambda_1\lambda_2^2\lambda_3\lambda_4}{S s_4 s_3 s_2^2 s_1}\right)\frac{s_2+s_3}{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)(\lambda_3+\lambda_4)}Z_2\,,\\
\Delta(\beta^{(3)}_{(1,3)},4)\Delta(\beta^{(4)}_{4},4) &= \left(\frac{\lambda_1\lambda_2^2\lambda_3\lambda_4}{S s_4 s_3 s_2^2 s_1}\right)\frac{s_2+s_4}{(\lambda_1+\lambda_4)(\lambda_2+\lambda_4)(\lambda_3+\lambda_4)}Z_3\,,
\end{aligned}$$

where

$$\begin{aligned}
Z_1 &= s_2 s_3(\lambda_1+\lambda_4)(s_4+\lambda_1)+s_2 s_4(\lambda_1+\lambda_3)(s_3+\lambda_1)+s_3 s_4(\lambda_1+\lambda_2)(s_2+\lambda_1),\\
Z_2 &= s_1 s_2(\lambda_3+\lambda_4)(s_4+\lambda_3)+s_1 s_4(\lambda_2+\lambda_3)(s_2+\lambda_3)+s_2 s_4(\lambda_1+\lambda_3)(s_1+\lambda_3),\\
Z_3 &= s_1 s_2(\lambda_3+\lambda_4)(s_3+\lambda_4)+s_1 s_3(\lambda_2+\lambda_4)(s_2+\lambda_4)+s_2 s_3(\lambda_1+\lambda_4)(s_1+\lambda_4),
\end{aligned}$$

and

$$S = \lambda_1+\lambda_2+\lambda_3+\lambda_4, \quad s_i = \sum_{\substack{j=1\\ j\neq i}}^{4}\lambda_j.$$

Then

(a) $\Delta(\beta^{(3)}_{(3,4)},4)\Delta(\beta^{(4)}_{1},4) \geq \Delta(\beta^{(3)}_{(1,4)},4)\Delta(\beta^{(4)}_{3},4)$ if and only if

$$\begin{aligned}
&(s_1+s_2)(\lambda_2+\lambda_3)(\lambda_3+\lambda_4)Z_1-(s_2+s_3)(\lambda_1+\lambda_2)(\lambda_1+\lambda_4)Z_2 =\\
&(\lambda_3-\lambda_1)\Big(f_1(\lambda_1,\lambda_2,\lambda_3,\lambda_4)+2\lambda_1\lambda_2\lambda_4^3(\lambda_4-\lambda_1)+2\lambda_2\lambda_3\lambda_4^3(\lambda_4-\lambda_2)+\\
&\qquad \lambda_4^3\left(\lambda_3^2\lambda_4-\lambda_2^3\right)+\lambda_4^4\left(\lambda_3\lambda_4-\lambda_2^2\right)\Big)\geq 0,
\end{aligned}$$

where $f_1(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$ is a polynomial with positive coefficients. Notice that the last four terms are positive because the λ_k's are increasing, even if 1 is replaced by 2.

(b) $\Delta(\beta^{(3)}_{(1,4)},4)\Delta(\beta^{(4)}_{3},4) \geq \Delta(\beta^{(3)}_{(1,3)},4)\Delta(\beta^{(4)}_{4},4)$ if and only if

$$\begin{aligned}
&(s_2+s_3)(\lambda_1+\lambda_4)(\lambda_2+\lambda_4)Z_2-(s_2+s_4)(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)Z_3 =\\
&(\lambda_4-\lambda_3)\left(f_2(\lambda_1,\lambda_2,\lambda_3,\lambda_4)+2\lambda_2\lambda_3(\lambda_3^4-\lambda_1^3\lambda_2)+2\lambda_1\lambda_4(\lambda_4^4-\lambda_1^2\lambda_2^2)+\right.\\
&\qquad \left.\lambda_1(\lambda_3^5-\lambda_1^3\lambda_2^2)+\lambda_1(\lambda_3^5-\lambda_1^2\lambda_2^3)\right)\geq 0,
\end{aligned}$$

where $f_2(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$ is a polynomial with positive coefficients. Using the ordering between the λ_k's, we can see that the last four terms are positive as before. □

Lemma 2.5.3. *Under the same assumptions as those in Lemma 2.5.2, then*

(a) $\Delta(\beta^{(3)}_{(2,u)},4)\Delta(\beta^{(4)}_1,4) \geq \Delta(\beta^{(3)}_{(1,2)},4)\Delta(\beta^{(4)}_u,4)$,

(b) $\Delta(\beta^{(3)}_{(1,u)},4)\Delta(\beta^4_2,4) \geq \Delta(\beta^{(3)}_{(1,2)},4)\Delta(\beta^{(4)}_u,4)$,

(c) if $\beta^{(3)}_{(1,u)} - \beta^{(4)}_2 < 0$, *then* $\Delta(\beta^{(3)}_{(2,u)},4)\Delta(\beta^{(4)}_1,4) \geq \Delta(\beta^{(3)}_{(1,u)},4)\Delta(\beta^{(4)}_2,4)$,

for $u = 3,4$.

PROOF. We give the proof only for the case $u = 3$. The case $u = 4$ is similar with 3 replaced by 4. After a few manipulations, we have

$$\begin{aligned}
\Delta(\beta^{(3)}_{(2,3)},4)\Delta(\beta^{(4)}_1,4) &= \left(\frac{\lambda_1\lambda_2\lambda_3\lambda_4^2}{S s_4^2 s_3 s_2 s_1}\right)\frac{s_1+s_4}{(\lambda_1+\lambda_2)(\lambda_1+\lambda_3)(\lambda_1+\lambda_4)}Y_1,\\
\Delta(\beta^{(3)}_{(1,3)},4)\Delta(\beta^{(4)}_2,4) &= \left(\frac{\lambda_1\lambda_2\lambda_3\lambda_4^2}{S s_4^2 s_3 s_2 s_1}\right)\frac{s_2+s_4}{(\lambda_1+\lambda_2)(\lambda_2+\lambda_3)(\lambda_2+\lambda_4)}Y_2,\\
\Delta(\beta^{(3)}_{(1,2)},4)\Delta(\beta^{(4)}_3,4) &= \left(\frac{\lambda_1\lambda_2\lambda_3\lambda_4^2}{S s_4^2 s_3 s_2 s_1}\right)\frac{s_3+s_4}{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)(\lambda_3+\lambda_4)}Y_3,
\end{aligned}$$

where

$$\begin{aligned}
Y_1 &= s_2 s_3(\lambda_1+\lambda_4)(s_4+\lambda_1) + s_2 s_4(\lambda_1+\lambda_3)(s_3+\lambda_1) + s_3 s_4(\lambda_1+\lambda_2)(s_2+\lambda_1),\\
Y_2 &= s_1 s_3(\lambda_2+\lambda_4)(s_4+\lambda_2) + s_1 s_4(\lambda_2+\lambda_3)(s_3+\lambda_2) + s_3 s_4(\lambda_1+\lambda_2)(s_1+\lambda_2),\\
Y_3 &= s_1 s_2(\lambda_3+\lambda_4)(s_4+\lambda_3) + s_1 s_4(\lambda_2+\lambda_3)(s_2+\lambda_3) + s_2 s_4(\lambda_1+\lambda_3)(s_1+\lambda_3).
\end{aligned}$$

(a) $\Delta(\beta^{(3)}_{(2,3)},4)\Delta(\beta^{(4)}_1,4) \geq \Delta(\beta^{(3)}_{(1,2)},4)\Delta(\beta^{(4)}_3,4)$ if and only if

$$\begin{gathered}
(s_1+s_4)(\lambda_2+\lambda_3)(\lambda_3+\lambda_4)Y_1 - (s_3+s_4)(\lambda_1+\lambda_2)(\lambda_1+\lambda_4)Y_3 =\\
(\lambda_3-\lambda_1)\big(f_3(\lambda_1,\lambda_2,\lambda_3,\lambda_4) + 2\lambda_2^2\lambda_4^2(\lambda_4-\lambda_2)(\lambda_3+\lambda_1) +\\
\lambda_4^2\left(\lambda_3^2\lambda_4^2-\lambda_2^4\right) + \lambda_4^3(\lambda_3\lambda_4^2-\lambda_2^3)\big) \geq 0,
\end{gathered}$$

where $f_3(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$ is a polynomial with positive coefficients. The last three terms are positive because the λ_k's are increasing, even if 3 is replaced by 4.

(b) $\Delta(\beta^{(3)}_{(1,3)},4)\Delta(\beta^{(4)}_2,4) \geq \Delta(\beta^{(3)}_{(1,2)},4)\Delta(\beta^{(4)}_3,4)$ if and only if

$$\begin{aligned}&(s_2+s_4)(\lambda_1+\lambda_3)(\lambda_3+\lambda_4)Y_2-(s_3+s_4)(\lambda_1+\lambda_2)(\lambda_2+\lambda_4)Y_3=\\&(\lambda_3-\lambda_2)\big(f_4(\lambda_1,\lambda_2,\lambda_3,\lambda_4)+\lambda_4^3(\lambda_3\lambda_4^2-\lambda_1^3)+\lambda_4^2(\lambda_2\lambda_4^3-\lambda_1^4)+\\&\qquad 2\lambda_1^2\lambda_4^2(\lambda_4-\lambda_1)(\lambda_2+\lambda_3)+\lambda_4^2(\lambda_2^4-\lambda_1^4)+\lambda_4^3(\lambda_2^3-\lambda_1^3)\big)\geq 0,\end{aligned}$$

where $f_4(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$ is a polynomial with positive coefficients. As earlier, we can see that the last five terms are also positive.

(c) $\Delta(\beta^{(3)}_{(2,3)},4)\Delta(\beta^{(4)}_1,4) \geq \Delta(\beta^{(3)}_{(1,3)},4)\Delta(\beta^{(4)}_2,4) \Leftrightarrow$

$$\begin{aligned}&(s_1+s_4)(\lambda_2+\lambda_3)(\lambda_2+\lambda_4)Y_1-(s_2+s_4)(\lambda_1+\lambda_3)(\lambda_1+\lambda_4)Y_2=\\&(\lambda_2-\lambda_1)\Big(f_5(\lambda_1,\lambda_2,\lambda_3,\lambda_4)+12\lambda_4^2\lambda_2^3(\lambda_3-\lambda_2)+13\lambda_4^2\lambda_2^2(\lambda_3\lambda_4-\lambda_2^2)+\\&\qquad\lambda_4^2\big((2\lambda_2-\lambda_1)^3-\lambda_3^3\big)(2\lambda_1+2\lambda_2+\lambda_4)+\lambda_4^2\big((2\lambda_2-\lambda_1)^4-\lambda_3^4\big)\Big)\geq 0,\end{aligned}$$

where $f_5(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$ is a polynomial with positive coefficients. Since the λ_k's are increasing, the last two terms in the penultimate row are positive and the final two terms are positive if $2\lambda_2-\lambda_1>\lambda_3$, i.e., if $\beta^{(3)}_{(1,u)}-\beta^{(4)}_2<0$ for $u=3$ or 4, which is assumed. □

CHAPTER 3

Sequential order statistics

Sequential order statistics are important in reliability studies such as a modification of k-out-of-n systems. As we pointed out, a k-out-of-n system is a system with n independent components which functions if and only if at least k of the n components are working. More generally, the failure of one component may influence the remaining components. Thus, a more flexible model for a k-out-of-n system should take the dependence structure into consideration. Thus, we shall suppose that, after each failure, the failure rates of the remaining, functioning components may change, so that the underlying failure rate of the remaining components is adjusted according to the number of preceding failures. For example, the breakdown of an aircraft's engine will increase the load put on the remaining engines, so that their operational lifetimes will tend to reduce. For dealing with this type of situation, sequential order statistics (SOS) were proposed by Kamps [34]. The SOS model is closely connected to several other models of ordered random variables and, in particular it unifies type II censored order statistics, k'th record values and k_n records from nonidentical distributions, see e.g. Cramer and Kamps [18].

Distributional and stochastic properties of ordinary order statistics have been studied extensively in the literature. Since SOS models unify various models of ordered random variables, it is interesting to study these characteristics but in relation with SOS. Cramer and Kamps [18] give an expression for the marginal distributions of SOS in terms of the so-called relevation transform (cf. Krakowski [46]). Zhuang and Hu [105] present some results on multivariate stochastic comparisons of SOS models and in particular, investigate conditions on the underlying distributions on which the SOS

models are based in order to obtain stochastic comparisons of SOS models in the multivariate likelihood ratio ($\leq_{lr}$), the dynamic multivariate hazard rate ($\leq_{dyn-hr}$) and the multivariate stochastic orders ($\leq_{st}$). See Shaked and Shanthikumar [88] for a review of this multivariate orders. The multivariate $\leq_{lr}$ and $\leq_{st}$ are closed under marginalization. Such a closure property is very useful to establish the univariate comparison result. The multivariate $\leq_{dyn-hr}$ is not closed under marginalization. Navarro and Burkschat [69] obtain some ordering properties for sequential k-out-of-n system based on SOS in order to study properties of the lifetimes of coherent systems based on SOS. They note that SOS are not necessarily ordered in hazard rate and likelihood ratio orderings.

In its general form the SOS model is linked with nonhomogeneous pure birth (NHPB) processes. In this field, there are several papers which study ageing notions of epoch times under conditions on the parameters of the NHPB process. Pellerey et al. [74] give conditions for the log-concavity of the density function of epoch times and inter-epoch times. Shaked et al. [89] highlight the relationship between l_∞-spherical densities and NHPB processes and provide applications to load sharing models, noting that studying the first n epoch times of a NHPB process is equivalent to studying the lifetimes of n components of a load sharing system. Results about multivariate stochastic comparisons of epoch times of two NHPB process have been given by Belzunce et al. [8]. They illustrate their results with applications to generalized Yule processes, load-sharing models, and minimal repairs in reliability theory. The intensity function $\lambda_i(t)$ of a NHPB process can be seen, in software reliability, as the software failures detection rate which depends on both the number of failures detected and time. When the failures detection rate not depend on time, then the NHPB process reduces to a homogeneous pure birth process (HPBP). Boland and Singh [11] investigated these processes as an approach to the Moranda geometric SR model.

The main motivation of this chapter is establish ageing properties and stochastic orderings of the epoch times of NHPB processes as an extension of the NHPP and HPBP processes, which are extensively used in software reliability modelling. Due to the relation between the epoch times of NHPB processes and the SOS, see Proposition 1.3.5 in Section 1.3, we study univariate stochastic comparisons of SOS and ageing notions, such as IHR, DHR and DRHR among others, which are not as strong as logconcavity of the density functions of a relevation counting process investigated by Pellerey et al. [74]. Note that if X is a random variable with logconcave density, then $\overline{F}$ is logconcave (IHR) and F is logconvex (DRHR).

The organization of this chapter is divided into seven sections. In Section 3.1, we first introduce some basic properties and definitions of SOS. We discuss, in Section

3.2, stochastic orderings between SOS and their underlying distribution function. In Section 3.3, we derive sufficient conditions under which the SOS are increasing hazard rate (IHR), increasing hazard rate average (IHRA) or decreasing hazard rate average (DHRA). We investigate, in Section 3.4, conditions on the underlying distribution functions on which the SOS are based, to obtain stochastic comparisons of SOS given the hazard rate and likelihood ratio orders. Examples of the underlying distributions, on which the SOS models are based, which satisfy these conditions are also given. Applications of the main results involving nonhomogeneous pure birth processes are given in Section 3.5. Section 3.6 is devoted to present a NHPB process approach to software reliability modelling. Finally, conclusions and possible extensions to this work are considered in Section 3.7. The results given in Sections 3.2-3.5 could be found in Torrado et al. [96].

3.1. DEFINITIONS AND PROPERTIES RELATED TO SOS

Cramer and Kamps [18] inspired the following definition of SOS given by Lenz [52].

Definition 3.1.1 (Lenz[52]). Let $G_1,\ldots,G_n$ be continuous distributions with $G_1^{-1}(1) \leq \ldots \leq G_n^{-1}(1)$ and let $X_{0,n}^* = -\infty$. Suppose that $U_i, i = 1,\ldots,n$ are independent random variables with $U_i \sim U(0,1)$. Then, the random variables

$$X_{i,n}^* = G_i^{-1}\left(1 - U_i \overline{G}_i(X_{i-1,n}^*)\right)$$

are called SOS based on $\{G_1,\ldots,G_n\}$.

Note that the above definition coincides with Definition 1.3.4 choosing

$$F_i(t) = 1 - \left(1 - G_i(t)\right)^{\frac{1}{n-i+1}}, \tag{3.1.1}$$

for $i = 1,\ldots,n$ (see Lenz[52]).

The marginal distribution functions $F_{*,1},\ldots,F_{*,n}$ of the SOS $X_{1,n}^*,\ldots,X_{n,n}^*$ based on $\{G_1,\ldots,G_n\}$ are given by,

$$\begin{aligned} F_{*,1}(t) &= G_1(t), \\ F_{*,i}(t) &= \begin{cases} F_{*,i-1}(t) - \displaystyle\int_{-\infty}^{t} \frac{\overline{G}_i(t)}{\overline{G}_i(z)}\, dF_{*,i-1}(z) & \text{if} \quad G_i(t) < 1, \\ 1 & \text{if} \quad G_i(t) = 1. \end{cases} \end{aligned} \tag{3.1.2}$$

From now on we shall assume that the distribution function of the i'th SOS is absolutely continuous with density function:

$$f_{*,i}(t) = h_i(t)\left(\overline{F}_{*,i}(t) - \overline{F}_{*,i-1}(t)\right), \tag{3.1.3}$$

where $h_i(t) = \frac{g_i(t)}{\overline{G}_i(t)}$, for all t.

Cramer and Kamps [18] noted that the corresponding distribution functions of SOS can be viewed as relevation transforms (Krakowski [46]). The relevation transform $\overline{F}\#\overline{G}$ of the survival functions $\overline{F}$ and $\overline{G}$ is defined by the Lebesgue-Stieltjes integral

$$\left(\overline{F}\#\overline{G}\right)(t) = \overline{F}(t) - \int_{-\infty}^{t} \frac{\overline{G}(t)}{\overline{G}(z)}\, d\overline{F}(z), \quad \text{for all } t.$$

Assuming that the supports of F and G are positive, then the relevation transform may be interpreted as the survival function of the time to failure of the second of two components when the second component with life distribution G is placed in service on the failure of the first component with life distribution F, assuming that the replacement component has the same age as the failed component (Lau and Prakasa Rao [51]). From (3.1.2), we have the representation

$$\overline{F}_{*,i}(t) = \overline{F}_{*,i-1}(t) - \int_{-\infty}^{t} \frac{\overline{G}_i(t)}{\overline{G}_i(z)}\, d\overline{F}_{*,i-1}(z), \quad \text{for all } t. \tag{3.1.4}$$

Hence, we can write the survival function of the i-th SOS as relevation transform

$$\overline{F}_{*,i} = \overline{F}_{*,i-1}\#\overline{G}_i.$$

Let us define,

$$A_i(t) = \int_{-\infty}^{t} \frac{1}{\overline{G}_i(z)}\, dF_{*,i-1}(z), \tag{3.1.5}$$

then, from (3.1.2) and (3.1.4) we have,

$$\overline{F}_{*,i}(t) = \overline{F}_{*,i-1}(t) + \overline{G}_i(t)A_i(t), \tag{3.1.6}$$

and

$$F_{*,i}(t) = F_{*,i-1}(t) - \overline{G}_i(t)A_i(t), \tag{3.1.7}$$

for $i = 2, \ldots, n$.

3.2. STOCHASTIC ORDERINGS BETWEEN SOS AND THEIR UNDERLYING CDF

In this section, we begin by reviewing some known results on stochastic orderings. We refer the reader to Subsection 1.4.1 for definitions of stochastic orderings used below. The following lemma, regarding the preservation of the hazard rate and reversed hazard rate orders under monotone increasing transformations, can be found in Keilson and Sumita [36].

Lemma 3.2.1 (Keilson and Sumita [36]). *Let X and Y be two random variables. If $X \leq_{hr} (\leq_{rh}) Y$, and if ϕ is any increasing function, then $\phi(X) \leq_{hr} (\leq_{rh}) \phi(Y)$.*

Shaked [87] also established the following relation between star ordering and dispersion ordering.

Lemma 3.2.2 (Shaked [87]). *Let X and Y be two non-negative random variables, then*

$$X \leq_* Y \Leftrightarrow \ln X \leq_{\text{disp}} \ln Y.$$

The next lemma due to Bartoszewicz [6] lists some relations between the dispersion order and other orders.

Lemma 3.2.3 (Bartoszewicz [6]). *Let X and Y be two random variables. Then,*

i) if X and Y are non-negative and $X \leq_{hr} Y$ and X or Y is DHR, then $X \leq_{\text{disp}} Y$;

ii) if $X \leq_{rh} Y$ and X or Y is IRHR, then $X \geq_{\text{disp}} Y$.

Now, we derive some preliminary results which are also of independent interest. In the next two lemmas, we show some stochastic orderings between SOS and their underlying distribution functions.

Lemma 3.2.4. *Let $X^*_{1,n},\ldots,X^*_{n,n}$ be SOS based on absolutely continuous distribution functions $\{G_1,\ldots,G_n\}$, then*

i) $G_i \leq_{hr} F_{*,i}$;

ii) $G_i \leq_{rh} F_{*,i}$;

iii) $G_i \leq_{lr} F_{*,i}$.

PROOF.

i) By definition, $G_i \leq_{hr} F_{*,i}$ if and only if $h_i(t) \geq h_{*,i}(t)$ for all t. From (3.1.3) we have

$$h_{*,i}(t) = h_i(t)\left(\frac{\overline{F}_{*,i}(t) - \overline{F}_{*,i-1}(t)}{\overline{F}_{*,i}(t)}\right).$$

Then, $h_i(t) \geq h_{*,i}(t) \Leftrightarrow \overline{F}_{*,i}(t) - \overline{F}_{*,i-1}(t) \leq \overline{F}_{*,i}(t)$.

ii) By definition, $G_i \leq_{rh} F_{*,i}$ if and only if $r_i(t) \leq r_{*,i}(t)$ for all t. First, we write the reversed hazard rate of the i'th SOS

$$r_{*,i}(t) = h_i(t)\left(\frac{\overline{F}_{*,i}(t) - \overline{F}_{*,i-1}(t)}{F_{*,i}(t)}\right) = r_i(t)\frac{G_i(t)}{\overline{G}_i(t)}\left(\frac{F_{*,i-1}(t) - F_{*,i}(t)}{F_{*,i}(t)}\right),$$

and then, we have that $r_i(t) \leq r_{*,i}(t)$ if and only if

$$\begin{aligned}
& 1 \leq \left(\frac{1-\overline{G}_i(t)}{\overline{G}_i(t)}\right)\left(\frac{F_{*,i-1}(t)-F_{*,i}(t)}{F_{*,i}(t)}\right) \\
\Leftrightarrow \quad & 0 \leq \frac{F_{*,i-1}(t)-F_{*,i}(t)-\overline{G}_i(t)F_{*,i-1}(t)}{\overline{G}_i(t)F_{*,i}(t)} \\
\Leftrightarrow \quad & F_{*,i}(t) = F_{*,i-1}(t) - \overline{G}_i(t)A_i(t) \leq G_i(t)F_{*,i-1}(t) \\
\Leftrightarrow \quad & F_{*,i-1}(t)\,(1-G_i(t)) \leq \overline{G}_i(t)A_i(t) \\
\Leftrightarrow \quad & F_{*,i-1}(t) \leq A_i(t) \\
\Leftrightarrow \quad & F_{*,i-1}(t) = \int_{-\infty}^{t} f_{*,i-1}(z)dz \leq \int_{-\infty}^{t} \frac{f_{*,i-1}(z)}{\overline{G}_i(z)}dz = A_i(t).
\end{aligned}$$

The last condition holds since $\overline{G}_i(t) \leq 1$ and from (3.1.5).

iii) By definition, $G_i \leq_{lr} F_{*,i}$ if and only if $f_{*,i}(t)/g_i(t)$ is increasing for all t. From (3.1.3) and (3.1.5) we have

$$f_{*,i}(t) = g_i(t)A_i(t) \Leftrightarrow \frac{f_{*,i}(t)}{g_i(t)} = A_i(t).$$

Clearly $A_i(t)$ is increasing, then $G_i \leq_{lr} F_{*,i}$ holds. □

Now, we present a connection between the SOS and their underlying distribution functions in the star ordering. First, let us define

$$u_i(t) = t \cdot h_i(t) \quad \text{and} \quad v_i(t) = t \cdot r_i(t).$$

Lemma 3.2.5. *Under the same assumptions as Lemma 3.2.4, if the support of G_i is non-negative for all i and*

i) if $u_i(t)$ is decreasing, then $G_i \leq_* F_{*,i}$ and

ii) if $v_i(t)$ is increasing, then $G_i \geq_* F_{*,i}$.

PROOF.

i) From Lemma 3.2.4(i) and Lemma 3.2.1 we have that $\ln G_i \leq_{hr} \ln F_{*,i}$. Now, the hazard rate of $\ln G_i$ is decreasing in t if and only if $u_i(t)$ is decreasing (see Theorem 2.3. in Kochar [41]). From Lemma 3.2.3(i), if $\ln G_i$ is DHR and $\ln G_i \leq_{hr} \ln F_{*,i}$, then $\ln G_i \leq_{\text{disp}} \ln F_{*,i}$. Finally, from Lemma 3.2.2 we have $\ln G_i \leq_{\text{disp}} \ln F_{*,i} \Leftrightarrow G_i \leq_* F_{*,i}$.

ii) From Lemma 3.2.4(ii) and Lemma 3.2.1 we have that $\ln G_i \leq_{rh} \ln F_{*,i}$. Now, it is easy to check that $v_i(t)$ is increasing if and only if the reversed hazard rate of $\ln G_i$ is increasing in t. From Lemma 3.2.3(ii), if $\ln G_i$ is IRHR and $\ln G_i \leq_{rh} \ln F_{*,i}$, then $\ln G_i \geq_{\text{disp}} \ln F_{*,i}$. Finally, from Lemma 3.2.2 we have $\ln G_i \geq_{\text{disp}} \ln F_{*,i} \Leftrightarrow G_i \geq_* F_{*,i}$.

□

It is worth noting that the condition that $u_i(t)$ is decreasing in Lemma 3.2.5(i) can be rewritten in the form $u_i'(t) = t \cdot h_i'(t) + h_i(t) \leq 0$. Therefore, it is clear that the condition that $h_i(t)$ be decreasing is a necessary but not sufficient condition for $u_i(t)$ to be decreasing. Similarly, the condition that $v_i(t)$ is increasing in Lemma 3.2.5(ii) can be rewritten as $v_i'(t) = t \cdot r_i'(t) + r_i(t) \geq 0$ and thus, it is clear that if $r_i(t)$ is increasing (i.e., X is IRHR) then $v_i(t)$ is also increasing (i.e., $\ln(X)$ is IRHR). However, the converse is not true as is illustrated by the following counterexample.

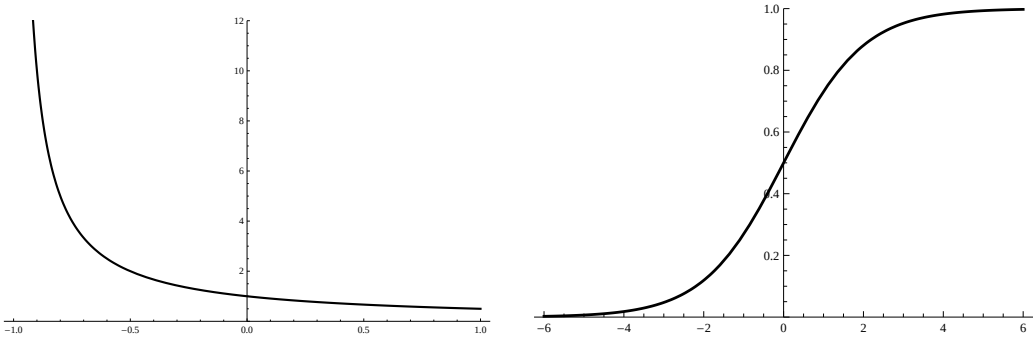
Counterexample 3.2.6 The reversed hazard rate of the uniform distribution on $[-1,1]$ is given by

$$r(t) = \frac{1}{1+t}, \quad t \in [-1,1].$$

As one can see from Figure 3.1, r is decreasing but the corresponding reversed hazard rate of the logarithm is

$$r_{\ln X}(t) = e^t r\left(e^t\right) = \frac{e^t}{1+e^t}, \quad \text{for all } t,$$

and it is easy to verify that $r_{\ln X}$ is increasing. ◁



(a) $X \sim \mathcal{U}(-1,1)$ (b) $\ln X$

Figure 3.1: Reversed hazard rate function

3.3. AGEING NOTIONS

Concepts of ageing, which were defined in Subsection 1.4.2, describe how a component or a system improves or deteriorate with age. Many classes of life distributions

can be categorized according to their ageing properties (see Marshall and Olkin [57]) . Below, we prove the following relationship between IRHR and IHR.

Lemma 3.3.1. *Let X be a random variable. Then,*

i) if X is IRHR, then its distribution function is convex,

ii) if X has a convex distribution function, then X is IHR.

PROOF. This proof comes from simple reasoning using the fact that $f(t) = r(t)F(t)$. Hence, it is clear that if $r(t)$ is increasing, then the density function is also increasing so that X has a convex distribution function. Now, if $f(t)$ is increasing, its hazard rate $h(t) = f(t)/\overline{F}(t)$ is increasing. Analyzing the following relation

$$h(t) = r(t)\frac{F(t)}{\overline{F}(t)},$$

one finds that if $r(t)$ is increasing, then $h(t)$ is also increasing. □

From Remark F.6. (pg. 179) in Marshall and Olkin [57], non-negative random variables cannot have distributions with increasing reversed hazard rate function. Thus, these random variables cannot satisfy Lemma 3.3.1(i).

Next, we study the preservation of some ageing notions through conditions on the underlying distribution function to the SOS. We begin by establishing a closure theorem for the convexity.

Theorem 3.3.2. *Let $X^*_{1,n},\ldots,X^*_{n,n}$ be SOS based on absolutely continuous distribution functions $\{G_1,\ldots,G_n\}$. If G_i is convex, then $F_{*,i}$ is convex for all i.*

PROOF. Note that, if G_i is convex, then g_i is increasing and from (3.1.3) we have

$$f_{*,i}(t) = g_i(t)A_i(t),$$

where $A_i(t)$ is defined in (3.1.5), and is also increasing, so that $F_{*,i}$ is convex. □

The previous theorem also provides a further result, namely that $F_{*,i}$ is IHR from Lemma 3.3.1(ii). It is also closely related to Theorem 4.5. in [18] where a particular choice of distribution functions, that is $G_i = 1-(1-F(t))^{\gamma_i}$, such that $\gamma_1,\ldots,\gamma_n > 0$ were considered and it was demonstrated that in this case, if F is IHR then $F_{*,i}$ is also IHR. It is noteworthy that, in this specific setup, SOS can be viewed as generalized order statistics and vice versa (Kamps [35], p. 56).

As noted in Shaked and Shanthikumar [88] the definition of the order $\leq_*$ is proper when the comparisons apply to distributions of non-negative random variables. This order can be used to characterize IHRA random variables as follows.

Lemma 3.3.3 (Shaked and Shanthikumar [88]). *Let* Exp *denote an exponential random variable and let X be a non-negative random variable. Then,*

$$X \text{ is IHRA (DHRA)} \Leftrightarrow X \leq_* (\geq_*)\text{Exp}.$$

By using this characterization we find conditions that imply that the SOS are IHRA and DHRA.

Theorem 3.3.4. *Let $X^*_{1,n},\ldots,X^*_{n,n}$ be SOS based on absolutely continuous distribution functions $\{G_1,\ldots,G_n\}$. If the support of G_i is non-negative, $u_i(t)$ is decreasing and G_i is DHRA, then $F_{*,i}$ is DHRA.*

PROOF. This proof depends on two lemmas. From Lemma 3.2.5(i) we know that $G_i \leq_* F_{*,i}$. Now, G_i is DHRA if and only if $G_i \geq_*$ Exp by the previous Lemma. Hence, Exp $\leq_* G_i \leq_* F_{*,i}$. □

Theorem 3.3.5. *Under the same assumptions as in Theorem 3.3.4, if $v_i(t)$ is increasing and G_i is IHRA, then $F_{*,i}$ is IHRA.*

PROOF. Now we apply the same method as in Theorem 3.3.4. By Lemma 3.2.5(ii) we know that $G_i \geq_* F_{*,i}$, and G_i is IHRA if and only if Exp $\geq_* G_i$ by the foregoing Lemma. Hence, Exp $\geq_* G_i \geq_* F_{*,i}$. □

The above theorems immediately lead to the following result because of the relation between SOS and generalized order statistics, when $G_i(t) = 1-(1-F(t))^{\gamma_i}$ and $\gamma_1,\ldots,\gamma_n$ are positive numbers.

Corollary 3.3.6. *Let $X^*_{1,n},\ldots,X^*_{n,n}$ be generalized order statistics based on an absolutely continuous distribution function F with hazard rate h and reversed hazard rate r and positive parameters $\gamma_1,\ldots,\gamma_n$.*

i) If $t\cdot h(t)$ is decreasing and F is DHRA, then $F_{*,i}$ is DHRA,

ii) If $t\cdot r(t)$ is increasing and F is IHRA, then $F_{*,i}$ is IHRA.

Below, we provide two examples in which the conditions of the previous theorems hold.

Example 3.3.7 **(Pareto distributions)** If the underlying distribution function of SOS is

$$G_i(t) = 1-\left(\frac{t}{c}\right)^{-\alpha_i}, \quad t \geq c \text{ and } \alpha_i > 1,$$

then, $u_i(t)$ is decreasing in t (see Example 2.2. in Kochar [41]). On the other hand, it is well-known that Pareto distribution is DHR. Therefore, G_i is DHRA and also concave. Hence the conclusions of Corollary 3.3.6(i) hold for this distribution. ◁

Example 3.3.8 **(Power function distributions)** If the underlying distribution function of SOS is

$$G_i(t) = \left(\frac{t}{c}\right)^{\alpha_i}, \quad 0 \le t \le c \text{ and } \alpha_i > 0,$$

then, the reversed hazard rate of the logarithm of G_i verifies the condition of Theorem 3.3.5 (see Example 2.1. in [41]). Also,

$$\frac{-\ln \overline{G}_i(t)}{t} = -\frac{1}{t}\ln\left(1-\left(\frac{t}{c}\right)^{\alpha_i}\right),$$

for $0 \le t < c$ is increasing if $\alpha_i \ge 1$. Furthermore, G_i is convex when $\alpha_i \ge 1$. Therefore, the conclusions of Theorems 3.3.2 and 3.3.5 hold. ◁

3.4. STOCHASTIC ORDERINGS

In this section, we investigate conditions on the underlying distribution functions on which the SOS are based, in order to obtain stochastic comparisons of SOS with various other univariate orders. Zhuang and Hu [105] presented some results on multivariate stochastic comparisons of SOS. They showed in their Theorem 3.7. that if the underlying distribution functions are ordered in the univariate hazard rate order, i.e., $G_1 \le_{hr} G_2 \le_{hr} \cdots \le_{hr} G_n$, then

$$\left(X^*_{1,n},\ldots,X^*_{n-1,n}\right) \le_{st} \left(X^*_{2,n},\ldots,X^*_{n,n}\right). \tag{3.4.8}$$

Since the usual multivariate stochastic order is closed under marginalization, we can get univariate comparisons of SOS from (3.4.8). However, in the univariate case, these results can be given without conditions, as Navarro and Burkschat [69] notice from Theorem 1.A.1. in [88], that is,

$$X^*_{1,n} \le_{st} X^*_{2,n} \le_{st} \cdots X^*_{n,n}. \tag{3.4.9}$$

They also obtain sufficient conditions to get the hazard rate order.

Theorem 3.4.1 (Navarro and Burkschat [69]). *Let $X^*_{1,n},\ldots,X^*_{n,n}$ be SOS based on $\{G_1,\ldots,G_n\}$. Let h_i denote the hazard rate function of G_i for $i = 1,\ldots,n$. If h_k/h_{k+1} is increasing for $k = 2,\ldots,i$, then*

$$X^*_{i,n} \le_{hr} X^*_{i+1,n} \quad \text{for } i = 1,\ldots,n-1.$$

We now proceed to stochastic comparisons of the first SOS and the others in the univariate hazard rate and likelihood ratio ordering.

Theorem 3.4.2. *Let $X^*_{1,n},\ldots,X^*_{n,n}$ be SOS based on absolutely continuous distribution functions $\{G_1,\ldots,G_n\}$, if $G_1 \leq_{hr(lr)} G_i$ for $i \geq 2$, then*

i) $X^*_{1,n} \leq_{hr} X^*_{i,n}$ and

ii) $X^*_{1,n} \leq_{lr} X^*_{i,n}$,

for $i = 2,\ldots,n$.

PROOF.

i) By definition we know that $X^*_{1,n} \leq_{hr} X^*_{i,n} \Leftrightarrow \overline{F}_{*,i}(t)/\overline{F}_{*,1}(t)$ is increasing in t. To do this we will use induction. It is immediately that $F_{*,1} \leq_{hr} F_{*,2}$ since from Lemma 3.2.4 we know that $G_2 \leq_{hr} F_{*,2}$ and by the assumptions $F_{*,1} = G_1 \leq_{hr} G_2$. We assume that $F_{*,1} \leq_{hr} F_{*,i-1}$, so we need to show that it is true for i. We get from (3.1.6) that
$$\frac{\overline{F}_{*,i}(t)}{\overline{F}_{*,1}(t)} = \frac{\overline{F}_{*,i-1}(t)}{\overline{F}_{*,1}(t)} + \frac{\overline{G}_i(t)A_i(t)}{\overline{F}_{*,1}(t)},$$
which is increasing in t since $A_i(t)$ and $\overline{G}_i(t)/\overline{F}_{*,1}(t)$ are increasing.

ii) In this case, $X^*_{1,n} \leq_{lr} X^*_{i,n} \Leftrightarrow f_{*,i}(t)/f_{*,1}(t)$ is increasing in t. We have, from (3.1.3), that
$$\frac{f_{*,i}(t)}{f_{*,1}(t)} = \frac{g_{*,i}(t)}{f_{*,1}(t)}A_i(t),$$
which is increasing in t since $A_i(t)$ and $g_i(t)/f_{*,1}(t)$ are increasing. □

Now, we discuss the likelihood ratio order. Navarro and Burkschat [69] show with the help of an example that the conditions in Theorem 3.4.1 for the hazard rate order are not sufficient conditions for the likelihood ratio order. Zhuang and Hu [105] give sufficient conditions for the likelihood ratio order.

Theorem 3.4.3 (Zhuang and Hu [105])**.** *Let $X^*_{1,n},\ldots,X^*_{n,n}$ be SOS based on $\{G_1,\ldots,G_n\}$. Let h_i denote the hazard rate function of G_i for $i = 1,\ldots,n$. If $G_i \leq_{lr} G_{i+1}$, and $h_{i+2} + h_i \geq 2h_{i+1}$ for $k = 1,\ldots,n-2$ when $n \geq 3$, then*
$$X^*_{i,n} \leq_{lr} X^*_{i+1,n} \quad \text{for } i = 1,\ldots,n-1.$$

We can now obtain alternative sufficient conditions for the likelihood ratio order of SOS. First, let us recall the definition of a TP2 function. A nonnegative function h of two variables, x and y, say, is called TP2 if $h(x',y)/h(x,y)$ is increasing in y whenever $x \leq x'$.

Lemma 3.4.4. *Let $X^*_{1,n},\ldots,X^*_{n,n}$ be SOS based on $\{G_1,\ldots,G_n\}$, if $\frac{g_{i-1}(t)}{g_i(t)}$ and $h_i(t)$ are TP2 in (i,t), and $G_{i-1} \leq_{hr} G_i$ for all i, then $A_i(t)$ is TP2 in (i,t) for $i = 3,\ldots,n$.*

PROOF. We will see, by induction on $i \geq 3$, that

$$A_i(t) = \int_{-\infty}^{t} \frac{1}{\overline{G}_i(z)} dF_{*,i-1}(z) = \int_{-\infty}^{t} \frac{g_{i-1}(z)}{g_i(z)} h_i(z) A_{i-1}(z) dz = \int_{-\infty}^{t} q_i(z) h_i(z) A_{i-1}(z) dz,$$

is TP2 in (i,t), where $q_i(z) = \frac{g_{i-1}(z)}{g_i(z)}$. For $i = 3$, by the assumptions, we have

$$\frac{q_3(t)}{q_2(t)} \frac{h_3(t)}{h_2(t)} A_2(t),$$

is increasing in t, which implies that $A_3(t)/A_2(t)$ is increasing in t. Let now $i \geq 4$. Again

$$\frac{q_i(t)\, h_i(t) A_{i-1}(t)}{q_{i-1}(t)\, h_{i-1}(t) A_{i-2}(t)},$$

is increasing in t, by the assumptions and by the induction hypothesis, which implies that $A_i(t)/A_{i-1}(t)$ is increasing in t. Hence, $A_i(t)$ is TP2 in (i,t). □

The following result gives conditions under which the SOS are comparable in the univariate likelihood ratio order.

Theorem 3.4.5. *Under the same assumptions than in Lemma 3.4.4, then*

$$X^*_{i-1,n} \leq_{lr} X^*_{i,n},$$

for $i = 3,\ldots,n$.

PROOF. By definition and from (3.1.3) we know that $X^*_{i-1,n} \leq_{lr} X^*_{i,n}$ iff

$$\frac{f_{*,i}(t)}{f_{*,i-1}(t)} = \frac{g_i(t) A_i(t)}{g_{i-1}(t) A_{i-1}(t)}, \tag{3.4.10}$$

is increasing in t. From the previous Lemma we know that $A_i(t)$ is TP2 in (i,t), and from Theorem 1.C.4(a) in [88] we get that $G_{i-1} \leq_{lr} G_i$, then, it follows that $f_{*,i}(t)/f_{*,i-1}(t)$ is increasing in t for $i = 3,\ldots,n$. □

Note that $g_{i-1}(t)/g_i(t)$ is TP2 in (i,t) can be written as

$$\frac{(g_{i-1}(t))^2}{g_i(t) g_{i-2}(t)}, \tag{3.4.11}$$

is increasing in t, and the condition $h_{i+2}+h_i \geq 2h_{i+1}$ in Theorem 3.4.3 is equivalent to

$$\frac{\left(\overline{G}_{i-1}(t)\right)^2}{\overline{G}_i(t)\overline{G}_{i-2}(t)}, \tag{3.4.12}$$

is increasing in t for $i = 1,\ldots,n-2$. Our previous result, Theorem 3.4.5, is equivalent to Theorem 3.4.3 in the sense that both have the same result and almost the same assumptions, except condition (3.4.11) and (3.4.12), respectively. Note that the condition (3.4.11) is useful when we have not an analytical expression of the survival functions.

When $G_i(t) = 1-(1-F(t))^{\gamma_i}$ for some distribution function F and γ_i are positive numbers for $i = 1,\ldots,n$, then $G_{i-1} \leq_{hr} G_i$ if and only if $\gamma_{i-1} \geq \gamma_i$, and the condition (3.4.11) holds if and only if $2\gamma_{i-1} \leq \gamma_i + \gamma_{i-2}$.

It is worth noting that the $(i-1)$'th SOS is not greater than the i'th SOS in the hazard rate and reversed hazard rate ordering as we will show in the following theorem. From (3.1.6) and (3.1.7), we get that

$$\frac{F_{*,i}(t)}{F_{*,i-1}(t)} = 1 - \frac{\overline{G}_i(t)A_i(t)}{F_{*,i-1}(t)}, \tag{3.4.13}$$

and

$$\frac{\overline{F}_{*,i}(t)}{\overline{F}_{*,i-1}(t)} = 1 + \frac{\overline{G}_i(t)A_i(t)}{\overline{F}_{*,i-1}(t)}, \tag{3.4.14}$$

for $i = 2,\ldots,n$.

Theorem 3.4.6. *Let $X^*_{1,n},\ldots,X^*_{n,n}$ be SOS based on absolutely continuous distribution functions $\{G_1,\ldots,G_n\}$, then*

i) $X^*_{i-1,n} \not\geq_{hr} X^*_{i,n}$ and

ii) $X^*_{i-1,n} \not\geq_{rh} X^*_{i,n}$,

for $i = 2,\ldots,n$.

PROOF.

i) Suppose that $X^*_{i-1,n} \geq_{hr} X^*_{i,n}$. By definition we know

$$X^*_{i-1,n} \geq_{hr} X^*_{i,n} \Leftrightarrow \frac{\overline{F}_{*,i}(t)}{\overline{F}_{*,i-1}(t)} \text{ is decreasing in } t,$$

and from (3.4.14), we have

$$X^*_{i-1,n} \geq_{hr} X^*_{i,n} \Leftrightarrow \frac{\overline{G}_i(t)A_i(t)}{\overline{F}_{*,i-1}(t)} \text{ is decreasing in } t.$$

Note that

$$\frac{\overline{G}_i(t)A_i(t)}{F_{*,i-1}(t)} = \frac{\overline{G}_i(t)A_i(t)}{\overline{F}_{*,i-1}(t)} \frac{\overline{F}_{*,i-1}(t)}{F_{*,i-1}(t)},$$

which is decreasing in t when $\frac{\overline{G}_i(t)A_i(t)}{\overline{F}_{*,i-1}(t)}$ is decreasing in t, since $\frac{\overline{F}_{*,i-1}(t)}{F_{*,i-1}(t)}$ is decreasing in t. Now, from (3.4.13)

$$\frac{\overline{G}_i(t)A_i(t)}{F_{*,i-1}(t)} \text{ is decreasing in } t \Leftrightarrow X^*_{i-1,n} \leq_{rh} X^*_{i,n},$$

i.e., if $X^*_{i-1,n} \geq_{hr} X^*_{i,n}$ then $X^*_{i-1,n} \leq_{rh} X^*_{i,n}$. Thus, $X^*_{i-1,n} =^{st} X^*_{i,n}$, which is a contradiction, since $X^*_{i-1,n} \leq_{st} X^*_{i,n}$ from (3.4.9). Hence $X^*_{i-1,n} \not\geq_{hr} X^*_{i,n}$.

ii) Suppose that $X^*_{i-1,n} \geq_{rh} X^*_{i,n}$. By definition we know

$$X^*_{i-1,n} \geq_{rh} X^*_{i,n} \Leftrightarrow \frac{F_{*,i}(t)}{F_{*,i-1}(t)} \text{ is decreasing in } t,$$

and from (3.4.13), we have

$$X^*_{i-1,n} \geq_{rh} X^*_{i,n} \Leftrightarrow \frac{\overline{G}_i(t)A_i(t)}{F_{*,i-1}(t)} \text{ is increasing in } t.$$

Note that

$$\frac{\overline{G}_i(t)A_i(t)}{\overline{F}_{*,i-1}(t)} = \frac{\overline{G}_i(t)A_i(t)}{F_{*,i-1}(t)} \frac{F_{*,i-1}(t)}{\overline{F}_{*,i-1}(t)},$$

which is increasing in t when $\frac{\overline{G}_i(t)A_i(t)}{F_{*,i-1}(t)}$ is increasing in t, since $\frac{F_{*,i-1}(t)}{\overline{F}_{*,i-1}(t)}$ is increasing in t. Now, from (3.4.14)

$$\frac{\overline{G}_i(t)A_i(t)}{\overline{F}_{*,i-1}(t)} \text{ is increasing in } t \Leftrightarrow X^*_{i-1,n} \leq_{hr} X^*_{i,n},$$

i.e., if $X^*_{i-1,n} \geq_{rh} X^*_{i,n}$ then $X^*_{i-1,n} \leq_{hr} X^*_{i,n}$. Thus, $X^*_{i-1,n} =^{st} X^*_{i,n}$, which is again a contradiction. Hence $X^*_{i-1,n} \not\geq_{rh} X^*_{i,n}$. □

A consequence of Theorem 3.4.6 is that $X^*_{i-1,n} \not\geq_{lr} X^*_{i,n}$ for $i = 2, \ldots, n$.

3.5. SOS AND NHBP PROCESSES

In this section, some applications of the main results in Sections 3.3 and 3.4 are presented. Specifically we give an application related to nonhomogeneous pure birth processes.

Nonhomogeneous pure birth processes are called relevation counting processes in Pellerey et al. [74], where some applications of them in reliability theory are described.

Another interpretation of these processes in reliability theory, by means of load sharing, is described in [89]. Recall the definition of a (NHPB) given in 1.2.2. A counting process $\{N(t), t \geq 0\}$ is a nonhomogeneous pure birth process with intensity functions $\{\lambda_i(t), i \geq 0\}$ and mean value functions $\{\Lambda_i(t), i \geq 0\}$, if the following hold

i) $N(t), t \geq 0$ has the Markov property;

ii) $P\{N(t+\Delta t) = i+1 | N(t) = i\} = \lambda_i(t)\Delta t + \circ(\Delta t)$ for $i \geq 1$;

iii) $P\{N(t+\Delta t) > i+1 | N(t) = i\} = \circ(\Delta t)$ for $i \geq 1$,

where the λ_i's are non-negative functions that satisfy

$$\int_t^\infty \lambda_i(x)dx = \infty, \qquad \text{for all } t \geq 0, \tag{3.5.15}$$

and

$$\Lambda_i(t) = \int_0^t \lambda_i(x)dx.$$

Condition (3.5.15) ensures that, with probability 1, the process has a jump after any time point t. When all the λ_i are identical, a nonhomogeneous pure birth process reduces to a nonhomogeneous Poisson process. We are especially interested in the coincidence (in distribution) of the epoch times of pure birth processes with certain models of ordered random variables such as record values, order statistics, generalized order statistics, Pfeifer record values (see Pfeifer [75]), and SOS. In a distributional theoretical sense, there is one-to-one correspondence between SOS and the first n epoch times of a NHPB process, which is stated in the following proposition.

Proposition 3.5.1 (Corollary 3.3.4. in Lenz [52])**.** *Let $G_1, \ldots, G_n$ be continuous distribution functions with $G_i(0) = 0$ and $G_i^{-1}(1) = c_i \in (0, \infty)$, $c_i \leq c_{i+1}$ and $X_{1,n}^*, \ldots, X_{n,n}^*$ the corresponding SOS. Let $\{N(t), t \geq 0\}$ be a NHPB process with mean value function $\Lambda_i(t)$ and denote the epoch times by $S_i, i = 1, \ldots, n$. Then S_i and $X_{i,n}^*$ coincide in distribution if and only if*

$$\Lambda_i(t) = -ln\,\overline{G}_i(t), \quad \text{for all } t \in [0, c_i).$$

Note that the above proposition coincides with Proposition 1.3.5 choosing F_i as in (3.1.1). Given this relationship, from Theorems 3.3.4, 3.3.5 and 3.4.2 and (3.4.9), it is possible to derive the following result.

Corollary 3.5.2. *Let $S_i, i \geq 1$ denote the epoch times of a NHPB process $\{N(t), t \geq 0\}$ with intensity functions $\lambda_i(t)$ and mean value function $\Lambda_i(t)$. Then:*

i) if $\Lambda_i(t)$ is antistarshaped and $t \cdot \lambda_i(t)$ is decreasing, then S_i is DHRA,

ii) *if* $\Lambda_i(t)$ *is starshaped and* $t \cdot \lambda_i(t)\left(e^{-\Lambda_i(t)}/1-e^{-\Lambda_i(t)}\right)$ *is increasing, then* S_i *is IHRA,*

iii) $S_{i-1} \leq_{st} S_i$, *for* $i = 2,\ldots,n$,

iv) *if* $\lambda_1(t) \geq \lambda_i(t)$ *for all* t *and for* $i = 2,\ldots,n$, *then* $S_1 \leq_{hr} S_i$, *for* $i = 2,\ldots,n$,

v) *if* $\lambda_1(t) \geq \lambda_i(t)$ *and* $\frac{\lambda_i(t)}{\lambda_1(t)}$ *is increasing in* t *for* $i = 2,\ldots,n$, *then* $S_1 \leq_{lr} S_i$, *for* $i = 2,\ldots,n$.

PROOF. Define

$$h_i(t) = \lambda_i(t) \qquad \text{for} \quad i = 1,\ldots,n.$$

Since (3.5.15) holds, $h_i(t)$ can be regarded as the hazard rate function of some distribution G_i. Let $X^*_{1,n},\ldots,X^*_{n,n}$ be the SOS based on distributions $\{G_1,\ldots,G_n\}$. Then, the result follows from Proposition 3.5.1 and Theorems 3.3.4–3.3.5 and 3.4.2 and equation (3.4.9). □

Note that, if we construct NHPB processes based on the distributions functions G_i of the Examples 3.3.7 and 3.3.8, then we get examples of NHPB processes which verify the conditions (i) and (ii) of the previous corollary, respectively.

3.6. APPLICATIONS IN SOFTWARE RELIABILITY MODELLING

In the context of software reliability, in our introductory chapter, we defined a software reliability model (SRM) as a mathematical tool to evaluate the software quantitatively. These stochastic models attempt to model either the times between successive failures of a piece of software or the number of failures in fixed time periods. The SRMs have been extensively developed in the literature. Most of them are based on stochastic counting processes, such as binomial process, pure birth process and non-homogeneous Poisson process (NHPP). We shall describe briefly two of them, specifically the model by Duane and a SRM based on a pure birth process. See Subsection 1.5.3 for a review on this models.

The Duane model [21] (DU) originally devised for hardware reliability model, is a infinite failures model. This model is a NHPP with the expected number of failures $\Lambda(t) = at^b$. The DU model could be stochastically represented as a Weibull process, in particular, this model is the counting process of the record values from a Weibull distribution.

Let us mention an homogeneous pure birth process (HPBP) for software reliability which is another variation of the Jelinski-Moranda model. This model is a birth process approach to the Moranda geometric SRM of Boland and Singh [11]. In this case,

the cumulative number of failures detected by time t is a HPBP with failure rates $\lambda_i = D \cdot k^i$, for $i = 0, 1, \ldots$, where $D > 0$ and $0 < k < 1$.

We present below a new model that incorporates all of these processes using the relationship between NHPB processes and SOS. We assume that, $N(t)$, the number of software failures in $(0,t]$ is a nonhomogeneous pure birth process. For any $t \geq 0$, define $N(0) = 0$ and $P_i(t) \equiv \Pr\{N(t) = i\}$ for $i = 1, 2, \ldots$. The Kolmogorov differential equations for a NHPB become (see, e.g., Parzen [73])

$$\begin{aligned} \frac{dP_1(t)}{dt} &= P_1'(t) = -\lambda_1(t) \cdot P_1(t), \\ \frac{dP_i(t)}{dt} &= P_i'(t) = -\lambda_i(t) \cdot P_i(t) + \lambda_{i-1}(t) \cdot P_{i-1}(t), \quad \text{for } i = 2, 3, \ldots \end{aligned}$$

and the probability generating function of $P_i(t)$ is

$$P(s,t) = \sum_{i=1}^{\infty} P_i(t) \cdot s^i, \quad \text{for any } t > 0 \text{ and } s \in [0,1].$$

Assumption 3.6.1. *In the following, we restrict ourselves to a particular choice of the distributions $G_1, \ldots, G_n$, namely*

$$G_i(t) = 1 - (1 - F(t))^{\gamma_i}, \quad 1 \leq i \leq n, \tag{3.6.16}$$

with some absolutely continuos and strictly increasing distribution function F and positive real numbers $\gamma_1, \ldots, \gamma_n$. Let f be the corresponding density function.

Although this setting seems to be very restrictive, many models of ordered random variables are included in this distribution theoretic sense (see e.g. Cramer and Kamps [18]) and also, this model is discussed in Cramer and Kamps [15, 16, 17].

From (3.6.16) we have

$$h_i(t) = \gamma_i h(t), \tag{3.6.17}$$

where $h(\cdot)$ is the hazard rate function of F. From Proposition 3.5.1 and (3.6.17), $N(t)$ is a NHPB process with birth rates

$$\lambda_i(t) = \gamma_i h(t), \quad i = 1, \ldots, n.$$

In order to reduce the model uncertainty to the parameters $\gamma_1, \ldots, \gamma_n$ we use specific distributions functions F. We start with a simple exponential family. Let F be given by

$$F(t) = 1 - e^{-\lambda g(t)}, \quad t \geq 0, \tag{3.6.18}$$

with $\lambda > 0$ unknown and some increasing and differentiable function g on $[0, \infty)$ satisfying $g(0) = 0$ and $\lim_{t \to \infty} g(t) = \infty$. The particular cases $g(t) = t^\beta$, $\beta > 0$, and $g(t) = \ln(t^\alpha)$,

$\alpha > 0$, correspond to standard Weibull (exponential, $\beta = 1$) and Pareto distributions, respectively. If $g(t) = t$, then the distribution function in (3.6.18) corresponds to an exponential distribution.

To apply this birth process to the reliability of a software program, we assume that software failures follow a nonhomogeneous pure birth process with birth rates

$$\lambda_i(t) = D \cdot k^i h(t), \quad i = 1, \ldots, n, \tag{3.6.19}$$

where $h(\cdot)$ is the hazard rate function of a exponential family of distributions. Note that if $g(t) = t$, then our model reduces to the HPBP model and if $g(t) = t^\beta$ and $k = 1$ then it reduces to de DU model.

We consider two-parameter Weibull distributions with $g(t) = t^\beta$ in (3.6.18) and we shall give an explicit expression for the mean value function and intensity function of the process $N(t)$, but first we obtain the probability generating function.

Lemma 3.6.2. *For any $t \geq 0$ and $s \in [0,1]$,*

$$P(s,t) = 1 - Dt\,h(t)(1-s) + \sum_{i=2}^{\infty} (-1)^i (Dt\,h(t))^i \cdot \left[\prod_{j=1}^{i} \left(1 - sk^{j-1}\right)\right] \cdot \left[\prod_{j=1}^{i-1} (j\beta + 1)\right]^{-1}. \tag{3.6.20}$$

PROOF. This proof is based on Boland and Singh [11].

$$\begin{aligned}
\frac{dP(s,t)}{dt} &= \sum_{i=1}^{\infty} \frac{dP_i(t)}{dt} s^i \\
&= -s\lambda_1(t)P_1(t) + \sum_{i=2}^{\infty} s^i \big(-\lambda_i(t)P_i(t) + \lambda_{i-1}(t)P_{i-1}(t) \big) \\
&= (s-1) \sum_{i=1}^{\infty} s^i \lambda_i(t) P_i(t) \\
&= -Dh(t)\,(1-s) \cdot P(ks,t).
\end{aligned}$$

This differential equation yields the integral equation

$$P(s,t) = 1 - Dh(t)(1-s) \int_0^t P(ks,t_1)dt_1.$$

Iterating this equation we have

$$\begin{aligned}
P(s,t) = 1 &- Dt\,h(t)(1-s) \\
&+ \sum_{i=2}^{n} (-1)^i (Dt\,h(t))^i \cdot \left[\prod_{j=1}^{i} \left(1 - sk^{j-1}\right)\right] \cdot \left[\prod_{j=1}^{i-1} (j\beta + 1)\right]^{-1} \\
&+ (-1)^{n+1} D^{n+1} h(t) \left[\prod_{j=1}^{n} \left(1 - sk^{j-1}\right)\right] \times \\
&\int_0^t h(t_1) \int_0^{t_1} h(t_2) \cdots \int_0^{t_n} P(k^{n+1}s, t_{n+1}) dt_{n+1} dt_n \cdots dt_1 .
\end{aligned} \tag{3.6.21}$$

One can check that the absolute value of the last term of (3.6.21) is $\leq$

$$D^{n+1}h(t)\int_0^t h(t_1)\int_0^{t_1} h(t_2)\cdots\int_0^{t_n} P(k^{n+1}s,t_{n+1})dt_{n+1}dt_n\cdots dt_1\,.$$

Integrating, we have that this term is equal to

$$\frac{\left(Dt\,h(t)\right)^{n+1}}{\prod_{i=1}^{n}(i\beta+1)} = \frac{\beta\left(Dt^{\beta}\right)^{n+1}}{\prod_{i=1}^{n}\left(i+\frac{1}{\beta}\right)} < \frac{\beta\left(Dt^{\beta}\right)^{n+1}}{n!}\,,$$

which tends to 0 as $n\to\infty$ for any fixed $t\geq 0$, since $\beta>0$ and $h(t)$ is the hazard rate of the Weibull distribution. □

The following theorem establishes an explicit expression for the mean value function and intensity function of the birth process.

Theorem 3.6.3. *Let $\{N(t), t\geq 0\}$ be the NHPB process with birth rates as in (3.6.19), then*

$$\Lambda(t) = Dt\,h(t) + \sum_{i=2}^{n}(-1)^{i-1}\left(Dt\,h(t)\right)^{i}\cdot\left[\prod_{j=1}^{i-1}\frac{\left(1-sk^{j}\right)}{(j\beta+1)}\right], \tag{3.6.22}$$

and

$$\lambda(t) = Dh(t)\cdot\left(\beta+\sum_{i=2}^{\infty}(-1)^{i-1}\left(Dt\,h(t)\right)^{i-1}j\beta\left[\prod_{j=1}^{i-1}\frac{\left(1-sk^{j}\right)}{(j\beta+1)}\right]\right). \tag{3.6.23}$$

PROOF. Differentiating (3.6.20) with respect to s for $0<s<1$, we have

$$\frac{dP(s,t)}{ds} = -\sum_{i=1}^{\infty}(-1)^{i}\left(Dt\,h(t)\right)^{i}\left(\prod_{j=0}^{i-1}\frac{1-sk^{j}}{j\beta+1}\right)\left(\sum_{j=0}^{i-1}\frac{k^{j}}{1-sk^{j}}\right).$$

Then, the mean value function is

$$E[N(t)] = \lim_{s\to 1}\frac{dP(s,t)}{ds} = Dt\,h(t) + \sum_{i=2}^{\infty}(-1)^{j-1}\left(Dt\,h(t)\right)^{i}\left(\prod_{j=1}^{i-1}\frac{1-k^{j}}{j\beta+1}\right).$$

Finally, we obtain (3.6.23) by differentiating (3.6.22). □

Note that (3.6.22) and (3.6.23) are similar to (1.5.19) and (1.5.20), the mean value function and the intensity function, respectively, of the SR model of Boland and Singh [11].

Next, we show expressions to estimate the parameters of the new model. Cramer and Kamps [15, 16, 17] are concerned with statistical inference for sequential k-out-of-n systems based on the distribution functions as defined in (3.6.16). In order to obtain the maximum likelihood estimators (MLE) of the parameters of the NHPB-based

software reliability model it is helpful to recognize that NHPB processes are closely connected to sequential order statistics. In particular, the joint density of $(X_1^*,\ldots,X_n^*)$ is given by

$$f_{X_1^*,\ldots,X_n^*}(\mathbf{x}) = n!\left[\prod_{i=1}^{n-1}\left(\frac{\overline{F}_i(x_i)}{\overline{F}_{i+1}(x_i)}\right)^{n-i} f_i(x_i)\right] f_n(x_n), \tag{3.6.24}$$

where $x_1 < x_2 < \cdots < x_n$ and $F_i = 1-(1-G_i)^{\frac{1}{n-i+1}}$ (see Lenz [52]). Choosing G_i as in (3.6.16) and F as in (3.6.18), then the joint density of $(X_1^*,\ldots,X_n^*)$ can be written as

$$f_{X_1^*,\ldots,X_n^*}(\mathbf{x}) = n!\left(\prod_{i=1}^{n}\frac{\gamma_i}{n-i+1}\right)\left(\prod_{i=1}^{n-1}\left(\overline{F}(x_i)\right)^{m_i} f_i(x_i)\right)\left(\overline{F}(x_n)\right)^{\frac{\gamma_i}{n-i+1}-1} f_n(x_n), \tag{3.6.25}$$

with $m_i = \gamma_i - \gamma_{i+1} - 1$. Recall from (1.1.2) the definition of the hazard rate function $h(t)$ and from (1.1.3) the definition of the cumulative hazard rate function $H(t)$. Then we have the logarithm of the likelihood function of the NHPB-SR model

$$\ln L(\gamma_i,F) = \ln n! + \sum_{i=1}^{n}\ln\left(\frac{\gamma_i}{n-i+1}\right) - \gamma_1 H(x_1) - \sum_{i=2}^{n}\gamma_i\Big(H(x_i)-H(x_{i-1})\Big) + \sum_{i=1}^{n}\ln\left(h(x_i)\right). \tag{3.6.26}$$

Choosing $\gamma_i = Dk^i$ with $D>0$ and $0<k\leq 1$, we get

$$\begin{aligned}\ln L(\gamma_i,F) = \ln n! + n\ln D + \frac{n(n+1)}{2}\ln k - \sum_{i=1}^{n}\ln(n-i+1) - DkH(x_1)\\ - \sum_{i=2}^{n}Dk^i\Big(H(x_i)-H(x_{i-1})\Big) + \sum_{i=1}^{n}\ln\left(h(x_i)\right).\end{aligned} \tag{3.6.27}$$

Maximizing equation (3.6.27) with respect to D and β, we have their maximum likelihood estimator

$$D = \frac{n}{kH(x_1) + \sum_{i=2}^{n}k^i\Big(H(x_i)-H(x_{i-1})\Big)}, \tag{3.6.28}$$

$$\sum_{i=1}^{n}\frac{d}{d\beta}\ln\left(h(x_i)\right) = D\left(k\frac{dH(x_1)}{d\beta} + \sum_{i=2}^{n}k^i\left(\frac{dH(x_i)}{d\beta} - \frac{dH(x_{i-1})}{d\beta}\right)\right). \tag{3.6.29}$$

Substituting equation (3.6.28) in (3.6.29), we get

$$\begin{aligned}\sum_{i=1}^{n}\frac{d}{d\beta}\ln\left(h(x_i)\right) &= n\,\frac{k\frac{dH(x_1)}{d\beta} + \sum_{i=2}^{n}k^i\left(\frac{dH(x_i)}{d\beta} - \frac{dH(x_{i-1})}{d\beta}\right)}{kH(x_1) + \sum_{i=2}^{n}k^i\Big(H(x_i)-H(x_{i-1})\Big)}\\ &= n\frac{d}{d\beta}\ln\left(kH(x_1) + \sum_{i=2}^{n}k^i\Big(H(x_i)-H(x_{i-1})\Big)\right).\end{aligned}$$

Thus, integrating the above equation, we have

$$\sum_{i=1}^{n} \ln\left(h(x_i)\right) = n \ln\left(kH(x_1) + \sum_{i=2}^{n} k^i \Big(H(x_i) - H(x_{i-1})\Big)\right). \tag{3.6.30}$$

Next, substituting equation (3.6.28) and (3.6.30) in (3.6.27), the logarithm of the likelihood function, and reducing the resultant expression, we have

$$\ln L(\gamma_i, F) = \ln n! + n \ln n + \frac{n(n+1)}{2} \ln k - \sum_{i=1}^{n} \ln(n-i+1) - n.$$

Then, it is easy to see that

$$\frac{d}{dk} \ln L(\gamma_i, F) = \frac{n(n+1)}{2} \frac{1}{k} > \frac{1}{k} > 0,$$

that is, the logarithm of the likelihood function in increasing in k. Thus, since $0 < k \leq 1$, the logarithm of the likelihood function reaches the maximum when $k = 1$. Therefore, the NHPB process reduces to a NHPP.

3.7. CONCLUDING REMARKS AND FURTHER WORK

In this chapter, we have studied ageing and ordering properties for the sequential order statistics which, as we have already mentioned, represent the lifetimes of a k-out-of-n systems with component lifetimes having a particular dependence model. We have investigated sufficient conditions on the underlying distribution functions on which the SOS are based, to obtain stochastic comparisons of SOS given the hazard rate and likelihood ratio orders. Also, we have derived conditions under which the SOS are increasing hazard rate, increasing hazard rate average or decreasing hazard rate average. Applications in software reliability involving nonhomogeneous pure birth processes are also given. Some results of this chapter are based on Torrado et al. [96].

There are many directions in which this research might be continued. Belzunce et al. [8] describe various conditions on the parameters of pairs of NHPB processes under which the corresponding epoch times or interepoch intervals are ordered in various senses. One possibility would be to identify conditions that enable one to compare spacings based on SOS from one sample. The other option would be to consider a Bayesian approach for incorporating additional prior information about the model parameters $\gamma_1, \ldots, \gamma_n$ when choosing the underlying distributions of SOS according to $G_i = 1 - (1-F)^{\gamma_i}$, for $1 \leq i \leq n$. To our knowledge, there exists only one reference in this topic, namely Burkschat et al. [13].

CHAPTER 4

Models with software metrics as covariate

As it was pointed out in Chapter 1, Section 1.5, a number of analytical models have been proposed for software reliability assessment. Most of these models are based on the assumption that the software is possibly imperfectly corrected after each failure or after various fixed time periods. In such cases, it is typically assumed that covariate information, in the form of software metrics such as code length and code complexity (McCabe [60]), will be generated each time the software is corrected. See Fenton and Pfleeger [22] for a good review of the main ideas.

Most software reliability models follow a non-Bayesian statistical point of view, in which the model parameters are considered to be fixed unknowns. When a large sample size of failure data is available, the parameters can be properly estimated using for example the maximum likelihood method (ML). However with a limited budget and little time for software development, it is difficult to obtain adequate sample data to accurately predict software reliability. In such situations, Bayesian approaches are used to determine parameters by incorporating prior information into the model, such as expert knowledge and past failure data. We shall here adopt a Bayesian approach to predict software failure data. For good reviews on this topic, see e.g. Singpurwalla and Wilson [91], Kuo and Yang [49] and Wiper [100].

The main purpose of this chapter is to present a novel software reliability model, which is a non-parametric regression model where software metrics data are used as covariate.

This chapter is organized as follows. We give, in Section 4.1, a brief description of software metrics and review software reliability models with software metrics in-

formation. In Section 4.2, we briefly outline some well known Bayesian methods in order to predict numbers of software failures and times between failures. Section 4.3 then describes our proposed model which can be seen as a non-parametric regression model based on Poisson failure counts and on exponentially distributed times between failures. In Subsection 4.3.4, we present the analysis of three real data sets and Section 4.4 finishes the chapter with some conclusions and possible extensions of our approach.

4.1. SOFTWARE METRICS INFORMATION

As we mentioned in Chapter 1, software development is a complicated process in which software faults are inserted in code by mistakes on the part of program developers. Some software programs are easy to modify, and therefore, they represent a small expense in developing computer systems. However, other programs are nearly impossible to modify without inserting multiple faults. Between these extremes lies a range of programs of intermediate complexity. Then, it would be reasonable to think that complex software programs are those likely to have a high fault count. Therefore, measures of software complexity can be used as good predictors of software quality. There are many software metrics, but here we only give a brief introduction, for more details on this topic see Fenton and Pfleeger [22] or Lyu [54].

The most used measure of source code program length is the *number of lines of code* (LOC). However without a careful definition of lines of code there will be many ways in which this measure may be calculated. To stress the fact that a line of code is actually a non-comment line we shall use the abbreviation (NCLOC). To derive NCLOC from a given software program you have to remove comments and blank lines. If CLOC is the number of comments and blank lines of program text, then it is useful to measure both NCLOC and CLOC and define total length, LOC, as

$$\text{LOC} = \text{NCLOC} + \text{CLOC}.$$

This way, we can immediately define some useful indirect measures like

$$\frac{\text{CLOC}}{\text{LOC}},$$

as a measure of the density of comments in a program.

It is also worth mentioning another commonly used software metrics developed by Halstead [29]. These metrics are now known as the *software science metrics*, and they are sensitive to program size but not to program control flow. That is, software programs with vastly different control flow structure can have identical Halstead metric values. Thus, Halstead's metric do not measure complexity due to control flow.

On the other hand, McCabe [60] developed a software metric, the *cyclomatic number*, which measures some aspects of control flow complexity. This metric is not necessarily related to program size, that is, software programs with vastly different LOC can have identical cyclomatic number. Thus, Halstead's metric and McCabe's metric measure two distinct program attributes. Each of these program attributes represents a source of variation underlying the measured complexity metrics.

4.1.1. Software reliability models using software metrics

As we mentioned, it is natural to suppose that software reliability can be related to software metrics information. However, the great majority of software reliability models do not take this covariate information into account. Important exceptions are Wiper and Rodríguez-Bernal [103] who modified the JM model to take account of software metrics information, Rinsaka et al. [82] who use a proportional hazard type approach, Ray et al. [81] who consider an approach based on estimating the number of bugs by a regression type model, and Wiper et al. [101] who use Bayesian neural networks.

Rinsaka et al. [82] and Ray et al. [81] consider a Type II software reliability model where the number of failures, $N_i = n_i$, detected by time t_i is modeled by a NHPP, for $i = 1, \ldots, M$. A natural model which incorporates software metrics is the following logistic regression model,

$$\begin{aligned} N_i \mid \lambda_i &\sim \mathcal{P}(\lambda_i) \\ \ln \lambda_i &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\,, \end{aligned}$$

where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})$ are the software metrics data. Wiper and Rodríguez-Bernal [103] consider a Bayesian approach to estimate the unknown parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^T$ of this model.

Rinsaka et al. [82] assume that the intensity function of the NHPP is

$$\lambda(t_i; \boldsymbol{\theta}, \boldsymbol{\beta} \mid \boldsymbol{x}_i) = \lambda_0(t_i; \boldsymbol{\theta}) \exp(\boldsymbol{x}_i \boldsymbol{\beta}),$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$ are coefficient parameters, and the function $\lambda_0(t_i; \boldsymbol{\theta})$ is called the baseline intensity function. They consider three kinds of baseline intensity functions: exponential $\lambda_0(t_i; a, b) = ab\exp(-bt)$, S-shaped $\lambda_0(t_i; a, b) = ab^2 \exp(-bt)$ and Rayleigh $\lambda_0(t_i; a, b) = atb^{-2} \exp(-t^2/2b^2)$.

Ray et al. [81] consider that the mean value function of the NHPP is

$$\Lambda(t_{i+1}) = r_j \Big(1 - \exp\left(-\beta_j (t_{j+1} - t_j)\right) \Big),$$

where r_j denotes the expected number of defects remaining at time t_j and the formulation for β_j consists of a loglinear model with

$$\ln \beta_j = \gamma_0 + \sum_{j=1}^{k} \gamma_j x_{ij} .$$

They use a Bayesian approach to estimate the parameters of the model.

Recently, Wiper et al. [101] develop a unified approach to Type I and Type II software reliability models in the presence of metrics information. They assume that $N_i \mid \lambda_i \sim \mathcal{P}(\lambda_i)$ for the Type II models and that $T_i \mid \lambda_i \sim \mathcal{E}(\lambda_i)$ for the Type I models, where T_i is the time between the i'th and the $(i-1)$'th software failures. In both cases, the failure rate λ_i follows a neural network model.

4.2. BAYESIAN INFERENCE

In this thesis we undertake Bayesian methods to predict software failures. The Bayesian approach may have an advantage over the non-Bayesian method when the parameters are unknown or difficult to estimate. It is well known that the main difference between the classical statistical theory and the Bayesian approach is that the latter considers parameters as random variables that characterized by a prior distribution. This *prior distribution* expresses the information available to the researcher before any data are involved in the statistical analysis. Interest lies in calculation of the *posterior distribution* $p(\boldsymbol{\theta} \mid \mathcal{D})$ of the parameters $\boldsymbol{\theta}$ given the observed data $\mathcal{D}$. According to the Bayes theorem, the posterior distribution can be written as

$$p(\boldsymbol{\theta} \mid \mathcal{D}) \propto \text{likelihood x prior} = p(\mathcal{D} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta}).$$

Specification of the prior distribution is important in Bayesian inference since it influences the posterior distribution. When no prior information of the parameters is available, we need to let the data "speak" for themselves. Such distributions are called *noninformative prior distribution*. Another nice property of a prior distribution is when the resulting posterior distribution is a member of the distributional family of the prior density. In this case, the prior distribution is called *conjugate prior distribution*.

This section provides a short introduction concerning the use of Markov chain Monte Carlo (MCMC) methods that are widely used in Bayesian inference and also, we review some well known Bayesian approaches to software reliability models.

4.2.1. Markov chain Monte Carlo algorithms

The focus of this subsection is on the most popular Markov chain simulation or MCMC methods. The most common application of these algorithms is the numerical

evaluation of high dimensional integrals. In particular, from the Bayesian point of view these methods are very useful in simulation from posterior distributions. There exits some software tools such as WinBUGS software, which use MCMC techniques to generate samples from posterior distribution of complicated models, providing an effective way to evaluate Bayesian models, see e.g., Ntzoufras [71].

MCMC methods are a class of algorithms for sampling from probability distributions, based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. A *Markov chain* is a stochastic process such that the distribution at sequence $i+1$ given all the preceding values depends only on the value of the previous sequence i.

Markov chain simulation is a general method based on drawing values of the parameters $\boldsymbol{\theta}$ from approximate distributions and then, correcting those values to better approximate the target posterior distribution $p(\boldsymbol{\theta} \mid \mathcal{D})$. The key to the method's success is that the approximate distributions are improved at each step in the simulation, in the sense of converging to the target distribution. Extensive details of the use of MCMC methods can be found in Robert and Casella [83].

The key to Markov chain simulation is to create a Markov process whose stationary distribution is the specified $p(\boldsymbol{\theta} \mid \mathcal{D})$ and run the simulation sufficiently long that the distribution of the current values is close enough to this stationary distribution. Once the simulation algorithm has been implemented and the simulation drawn, it is necessary to check the convergence of the simulated sequences. In this thesis, we will apply two main types of MCMC algorithm: the Metropolis-Hastings algorithm and the Gibbs sampler.

The result of the *Metropolis-Hastings algorithm* will be the simulation of an objective or target density. In the Bayesian framework, this target density is the posterior distribution, from which we wish to generate a sample of size M. The Metropolis-Hastings algorithm can be described by the following iterative steps, where $\boldsymbol{\theta}^{(i)}$ are the parameters of interest of generated values in i iteration of the algorithm:

1. Set initial values $\boldsymbol{\theta}^{(0)}$.

2. For $i = 1, \ldots, M$, repeat the following steps:

   a. Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(i-1)}$.

   b. Generate new candidate parameter values $\boldsymbol{\theta}'$ from a proposal distribution $q(\boldsymbol{\theta}' \mid \boldsymbol{\theta})$.

   c. Calculate

$$\alpha = \min\left\{1, \frac{p(\boldsymbol{\theta}' \mid \mathcal{D})\, q(\boldsymbol{\theta} \mid \boldsymbol{\theta}')}{p(\boldsymbol{\theta} \mid \mathcal{D})\, q(\boldsymbol{\theta}' \mid \boldsymbol{\theta})}\right\}.$$

d. Update $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}'$ with probability α, otherwise set $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}$.

The Metropolis-Hastings algorithm will converge to its equilibrium distribution regardless of whatever proposal distribution q is selected. Another important characteristic of this algorithm is that we do not need to evaluate the normalizing constant $p(\mathcal{D})$ involved in $p(\boldsymbol{\theta} \mid \mathcal{D})$ since it cancels out in α.

The *Gibbs sampler* is a special case of Metropolis-Hastings sampling where the acceptance probability, α, is equal to 1, and therefore the generated parameter is accepted in all iterations. One advantage of the Gibbs sampler is that, in each step, random values must be generated from unidimensional distributions for which a wide variety of computational tools exists. Frequently, these conditional distributions have a known form and, thus, random numbers can be easily simulated using standard functions.

The algorithm can be summarized by the following steps:

1. Set initial values $\boldsymbol{\theta}^{(0)}$.

2. For $i = 1, \ldots, M$, repeat the following steps:

   a. Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(i-1)}$.

   b. For $j = 1, \ldots, k$, update θ_j from $\theta_j \sim p(\theta_j \mid \theta_1, \ldots \theta_{j-1}, \theta_{j+1}, \ldots, \theta_k, \mathcal{D})$.

   c. Set $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}$ and save it as the generated set of values at i iteration of the algorithm.

4.2.2. Bayesian approaches to software reliability modelling

Due to the large number of software reliability models which were developed from the Jelinski-Moranda model, some authors have attempted to unify these models from different perspectives. The advantage of unification is the availability of a common structure under which the problem of reliability growth or decay can be studied. A unifying perspective on the many software reliability models can hopefully simplify the task of model selection that user faces. One of the attempts at unifying the software reliability models is adopting a Bayesian point of view (see, e.g. Langberg and Singpurwalla [50]). Here, we review Bayesian approach to Type I and Type II SR models, some of them are Bayesian versions of the SR models presented in Section 1.5.

Recall that the Jelinski-Moranda (JM) model, which we defined in Chapter 1, supposes that the distribution of the i'th interfailure times is given by

$$T_i \mid N, \phi \sim \mathcal{E}\big((N-i+1)\phi\big), \tag{4.2.1}$$

an exponential distribution with mean $1/\big((N-i+1)\phi\big)$.

The first Bayesian software reliability model is due to Littlewood and Verrall (LV) [53]. They proposed a model to relax the assumption of perfect repair in the JM model. Here, the $(N-i+1)\phi$ of (4.2.1) is replaced by an unknown parameter λ_i. They consider $\lambda_i \sim \mathcal{G}(\alpha, \psi(i))$, a Gamma distribution as the prior of λ_i, where $\psi(i)$ is an increasing function of i to reflect that repairs tend to make the software more reliable. Littlewood and Verrall [53] considered $\psi(i)$ to be completely specified, and also, methods of estimation.

A particular form, $\psi(i) = \beta_0 + \beta_1 i$, was proposed by Mazzuchi and Soyer [59]. This form ensures that $\frac{\alpha}{\psi(i)}$, the expected value of λ_i, decreases in i. Both α and $\psi(i)$ are treated as unknown. The prior distributions are as following,

$$\begin{aligned} \alpha &\sim \mathcal{U}[0, w], \quad \text{an Uniform distribution with } w > 0, \\ \beta_1 &\sim \mathcal{G}(u_1, v_1), \quad \text{a Gamma distribution independent of } \alpha, \text{ and} \\ \beta_0 \mid \beta_1 &\sim \mathcal{SG}(u_0, v_0; \beta_1), \end{aligned}$$

a shifted Gamma distribution with β_1 being the extend of the shift. This model has been extended by Kuo and Yang [48] who take $\psi(i)$ to be a polynomial of degree k, and by Soyer [92] who lets λ_i have expectation αi^{β}.

Bayesian inference for the JM model has also been examined in, e.g., Meinhold and Singpurwalla [62] and Wiper et al. [102], among others. Here we just give a brief introduction. We shall consider the following prior distribution for N and ϕ,

$$\begin{aligned} N &\sim \mathcal{P}(\lambda), \quad \text{a Poisson distribution with mean } \lambda, \text{ and} \\ \phi &\sim \mathcal{G}(\alpha, \beta), \quad \text{a Gamma distribution.} \end{aligned}$$

Given the observation of n failures, then the conditional posterior distribution are,

$$\begin{aligned} N - n \mid \mathcal{D} &\sim \mathcal{P}\big(\lambda \exp(-n\bar{t}\phi)\big), \quad \text{where } \bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i, \\ \phi \mid \mathcal{D} &\sim \mathcal{G}\left(\alpha + n, \beta + \sum_{i=1}^{n}(N-i+1)t_i\right). \end{aligned}$$

It is useful to note that even though N and ϕ were a priori independent, once the data $\mathcal{D}$ are at hand, they are a posteriori dependent. This is to be expected because posterior inference for both parameter is based on the same set of data.

In Section 1.5.2 we introduced a Type II model by Goel and Okumoto (GO), called a *time dependent error detection model*. Recall that software failures are encountered as a nonhomogeneous Poisson process (NHPP), $\{N(t), t \geq 0\}$, with mean value function $\Lambda(t) = a(1 - e^{-bt})$, where a and b are unknown parameters. McDaid and Wilson

[61] proposed a Bayesian analysis of the GO model by assuming independent gamma priors on a and b as following,

$$\begin{aligned} N(t) \mid a,b &\sim \mathcal{P}\big(a(1-e^{-bt})\big), \\ a &\sim \mathcal{G}(\alpha_a,\beta_a), \\ b &\sim \mathcal{G}(\alpha_b,\beta_b). \end{aligned}$$

A model which provides an unification of the GO model and LV model is due to Rodrigues [84]. He formulated the following hierarchical Bayes SR model.

$$\begin{aligned} T_i \mid N,\phi &\sim \mathcal{E}\big((N-i+1)\phi\big), \\ N \mid \mu &\sim \mathcal{P}(\mu), \\ \mu &\sim \mathcal{G}(\alpha_\mu,\beta_\mu), \\ \phi &\sim \mathcal{G}(\alpha_\phi,\beta_\phi). \end{aligned}$$

Assume that μ and ϕ are known values, then we have the GO model. If N and μ are known values, then we get the LV model, and when μ is a known value, then the model by Rodrigues [84] is a combination of the GO and LV models.

Many other Bayesian SR models have been developed. Kuo and Yang [49] gave the Gibbs algorithms incorporating Metropolis-Hasting methods to 8 kinds of NHPP-based SR models. More recently, Ruggeri and Soyer [85] propose a model based on a hidden Markov chains, which assumes that times between failures are exponentially distributed with parameters depending on an unknown latent state variable which, in turn, evolves as a Markov chain. Hirata et al. [30] developed a Java based prototype tool for Bayesian estimation in NHPP -based SR models.

4.3. A Bayesian approach to SRMs using covariates

Often, in real software development scenarios, members of the developing time or software users may have useful prior knowledge about the quality of the software and then, via a Bayesian approach, this knowledge could be used to improve the predictions of the proposed model. Even when real prior knowledge is not available and relatively uninformative priors are proposed, then in a predictive system such as in the software reliability scenario, Bayesian predictions automatically account for any parameter uncertainty. However, classical approaches based on substituting parameters by plug in estimates often underestimate predictive uncertainty. For more comments of the advantages of the Bayesian approach in software reliability, see e.g. Singpurwalla and Wilson [91].

In this section, we shall develop an alternative approach to both Type I and Type II software reliability models, where the failure rate at a given time period is estimated using Bayesian non-parametric regression techniques based on Gaussian processes, after the software has been corrected. In the following subsection, we give a brief review on Gaussian processes. We also consider the software metrics for the current version of the software as the covariate in our Bayesian non-parametric regression model. This approach may be thought of as an extension of the work of Ray et al. [81] which generalizes this earlier, parametric regression based approach by using a nonparametric regression model. Although Bayesian models for software reliability have been investigated in the literature, as we showed in Section 4.2.2, to the best of our knowledge, these models do not incorporate any covariate information.

4.3.1. Gaussian processes

Gaussian process models have recently been used in Bayesian approaches to regression, classification and other areas, see e.g., Rasmussen and Williams [80]. Formally, a Gaussian process is defined as following.

Definition 4.3.1. A *Gaussian process* (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution.

A GP is a generalization of the Gaussian probability distribution. Just as a Gaussian distribution is fully specified by its mean and covariance matrix, a GP is specified by a mean and a covariance function. We define the mean function $m(\boldsymbol{x})$ and the covariance function $C(f(\boldsymbol{x}), f(\boldsymbol{x}'))$ of a real process $f(\boldsymbol{x})$ as

$$\begin{aligned} m(\boldsymbol{x}) &= \mathbb{E}[f(\boldsymbol{x})], \\ C(f(\boldsymbol{x}), f(\boldsymbol{x}')) &= \mathbb{E}\left[(f(\boldsymbol{x}) - m(\boldsymbol{x}))\left(f(\boldsymbol{x}') - m(\boldsymbol{x}')\right)\right]. \end{aligned}$$

As we assume the mean function is defined to be zero, the covariance function is the crucial ingredient in a GP model. There exist different covariance functions, as the *dot product* covariance function which depends only on $\boldsymbol{x}$ and $\boldsymbol{x}'$ through $\boldsymbol{x} \cdot \boldsymbol{x}'$, or the *inhomogeneous polynomial kernel*, $C(f(\boldsymbol{x}), f(\boldsymbol{x}')) = (\sigma^2 + \boldsymbol{x} \cdot \boldsymbol{x}')^r$, where r is a positive integer, among others. In addition, the covariance function can be the sum or product or linear combination of different covariance functions. The only technical restriction on the covariance function is that it must be positive semidefinite. Following Rasmussen and Williams [80], we define the *squared exponential covariance function* as

$$C(f(\boldsymbol{x}), f(\boldsymbol{x}') \mid \boldsymbol{\theta}) = \eta^2 \exp\left\{-\frac{1}{2}\sum_{j=1}^{k} \rho_j^{-2}\left(x_j - x_j'\right)^2\right\}, \tag{4.3.2}$$

where $\boldsymbol{\theta} = (\rho_1^2, \ldots, \rho_k^2, \eta^2)$ is the unknown parameter set, i.e., the *hyperparameter* set. Observe that ρ_j is a length scale parameter and the variance η^2, called the scaling parameter, is the overall vertical scale of variation of the latent value.

Note that for n observed data points, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, the mean vector $\boldsymbol{m}$ will be an n-element column vector, and the *covariance matrix* $C(\boldsymbol{\theta})$ will be an $n \times n$ matrix with elements as given in (4.3.2).

GPs are used in regression and classification problems. Next, we will review briefly the Bayesian analysis of the standard regression model. Given a set $\mathcal{D} = \{(\boldsymbol{x}_i, y_i) : i = 1, \ldots, n\}$ of n observations, where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})^T$ denotes an input vector (covariates) of dimension k and $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ denotes a scalar output or target (dependent variable), the regression model is defined as

$$y_i = f(\boldsymbol{x}_i) + \varepsilon_i,$$

where ε_i follows an independent, identically distributed Gaussian distribution with zero mean and variance σ^2. Usually, one assume that the prior distribution, $p(\boldsymbol{f} \mid \boldsymbol{\theta})$, on $\boldsymbol{f} = (f_1, \ldots, f_n)^T$ is a GP, where $f_i = f(\boldsymbol{x}_i)$, and we will write the GP as $\boldsymbol{f} \mid \boldsymbol{\theta} \sim \mathcal{GP}(m(\boldsymbol{x}), C(\boldsymbol{\theta}))$. In a regression problem with additive noise the likelihood is straightforwardly obtained from the noise model of ε_i, thus, the likelihood is Gaussian. Then the posterior distribution on $\boldsymbol{f}$, $p(\boldsymbol{f} \mid \mathcal{D}, \boldsymbol{\theta}) \propto p(\mathcal{D} \mid \boldsymbol{f}) \times p(\boldsymbol{f} \mid \boldsymbol{\theta})$, is also a GP.

When the posterior distribution cannot be computed analytically, some approximation must be employed to obtain an approximate posterior. Several different techniques have been proposed to overcome this obstacle, such as Laplace approximation, expectation propagation algorithm and MCMC sampling. For a deep study on this topic, we refer the reader to an excellent book by Rasmussen and Williams [80].

4.3.2. Model description

In this subsection we present a new approach to both Type I and Type II software reliability models (see Section 1.5). Our model is a hierarchical non-parametric regression model based on exponential interfailure times or Poisson failure counts where the rates are modeled as Gaussian processes where software metrics data are used as inputs.

We consider a regression problem where we have a data set $\mathcal{D}$ of M scalar observations with an arbitrary distribution with parameter λ_i and that the software being analyzed is possibly imperfectly corrected after each period. If we assume that we observe the times between successive M failures, say $T_1 = t_1, \ldots, T_M = t_M$, then, in this case, $\mathcal{D} = \{t_i : i = 1, \ldots, M\}$. We might also assume that interfailure times are exponentially

distributed, that is,

$$S_i \mid \lambda_i \sim \mathcal{E}(\lambda_i).$$

When we assume that we observe the numbers of failures, say $N_1 = n_1, \ldots, N_M = n_M$ in M time periods of length $L_1, \ldots, L_M$ respectively, then $\mathcal{D} = \{n_i : i = 1, \ldots, M\}$. We also assume that the numbers of failures follows a Poisson distribution, that is, for $i = 1, \ldots, M$, we have

$$N_i \mid \lambda_i \sim \mathcal{P}(L_i \lambda_i).$$

As part of the correction procedure, we shall suppose that after the $(i-1)$'th failure the software is possibly imperfectly corrected and software metrics, say $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})$ are generated for $i = 1, \ldots, M$. Such metrics may reflect both characteristics of the code such as number of lines or also measures of the amount of work undertaken on correction such as many hours or costs. Thus, it is reasonable to suppose that changes in the quality of the code will be reflected in changes in the values of the software metrics.

In both cases, the rate, λ_i can be modeled as a function of the software metrics, $\boldsymbol{x}_i$, available after the last correction as

$$\ln \lambda_i \mid f_i = f(\boldsymbol{x}_i) + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2)$ and $f : \Re^k \to \Re$ can take different forms.

Now, the most important problem to consider is how to model the unknown function, f. One possibility is to assume that f is a linear function of the software metrics, say

$$f(\boldsymbol{x}_i) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij},$$

see Subsection 4.1.1, but there is quite a lot of evidence to illustrate that the relationship between software quality and software metrics is often highly non-linear and therefore, it seems preferable to use a more general, fully nonparametric model. One possibility would be to use classical, nonparametric regression techniques, see e.g. Fox [24], but here we prefer to use a Bayesian approach, as outlined below.

We shall follow a popular approach in the Bayesian literature of modeling f using Gaussian process (GP) priors, see Subsection 4.3.1. A GP prior supposes that for any finite set of points, say $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$, the joint distribution of $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_M)$ is normal or Gaussian distributed with zero mean and a squared exponential covariance function defined as in (4.3.2). Here we shall write

$$\boldsymbol{f} \mid \boldsymbol{\theta} \sim \mathcal{GP}(0, C(\boldsymbol{\theta})),$$

where $\boldsymbol{\theta}$ are the parameters characterizing the process and $\boldsymbol{f}$ denotes a set of latent variables in a vector. This implies that the probability density of $\ln(\boldsymbol{\lambda}) = \big(\ln\lambda_1,\ldots,\ln\lambda_M\big)^T$, given the parameters, is a factorized Gaussian, $\ln(\boldsymbol{\lambda}) \mid \boldsymbol{f},\sigma^2 \sim \mathcal{N}(\boldsymbol{f},\sigma^2\boldsymbol{I})$ and therefore, $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_M)^T$, given the parameters, is a lognormal distribution

$$p(\boldsymbol{\lambda} \mid \boldsymbol{f},\sigma^2) \sim \mathcal{LN}(\boldsymbol{f},\sigma^2\boldsymbol{I}), \tag{4.3.3}$$

where $\boldsymbol{I}$ is the identity matrix. The basic, Bayesian model is then completed by defining prior distributions for the error variance, σ^2 and for the GP parameters, $\boldsymbol{\theta}$. As is typical in such problems, in the lack of strong prior information, we shall assume independent, proper but relatively uninformative inverse gamma ($\mathcal{IG}$) priors for σ^2, η^2 and ρ_j^2 for $j = 1,\ldots,k$. The family of inverse gamma densities is versatile enough to incorporate both increasing and decreasing failure rates.

Therefore, we have an approach to both Type I and Type II SR models which can be summarize, respectively, as following.

$$\begin{aligned} T_i \mid \lambda_i &\sim \mathcal{E}(\lambda_i), \\ \ln\lambda_i \mid f_i &= f(\boldsymbol{x}_i)+\varepsilon_i, \\ \boldsymbol{f} \mid \boldsymbol{\theta} &\sim \mathcal{GP}(0,C(\boldsymbol{\theta})). \end{aligned} \qquad \begin{aligned} N_i \mid \lambda_i &\sim \mathcal{P}(L_i\lambda_i), \\ \ln\lambda_i \mid f_i &= f(\boldsymbol{x}_i)+\varepsilon_i, \\ \boldsymbol{f} \mid \boldsymbol{\theta} &\sim \mathcal{GP}(0,C(\boldsymbol{\theta})). \end{aligned}$$

Next, we can theoretically evaluate the posterior distributions using Bayes theorem. The joint posterior for $\boldsymbol{\lambda}$,$\boldsymbol{f}$ and σ^2 given the observed data, $\mathcal{D}$, is

$$\begin{aligned} p(\boldsymbol{\lambda},\boldsymbol{f},\sigma^2 \mid \mathcal{D}) &\propto \text{likelihood} \times \text{prior} \\ &= p(\mathcal{D} \mid \boldsymbol{\lambda}) \times p(\boldsymbol{\lambda},\boldsymbol{f},\sigma^2) \\ &= p(\mathcal{D} \mid \boldsymbol{\lambda}) \times \Big(p(\boldsymbol{\lambda} \mid \boldsymbol{f},\sigma^2)\cdot p(\boldsymbol{f},\sigma^2)\Big). \end{aligned}$$

Unfortunately simple closed form expressions for the integration constants of the posterior parameter distributions are not available. Thus, we need to use numerical integration techniques or simulation methods to generate samples from these posterior distributions. Note that by decomposition:

$$\begin{aligned} p(\boldsymbol{\lambda},\boldsymbol{f},\sigma^2 \mid \mathcal{D}) &= p(\boldsymbol{\lambda} \mid \boldsymbol{f},\sigma^2,\mathcal{D})\cdot p(\boldsymbol{f},\sigma^2 \mid \mathcal{D}) \\ &= p(\boldsymbol{\lambda} \mid \boldsymbol{f},\sigma^2,\mathcal{D})\cdot p(\boldsymbol{f} \mid \sigma^2,\mathcal{D})\cdot p(\sigma^2 \mid \mathcal{D}), \end{aligned} \tag{4.3.4}$$

where

$$p(\boldsymbol{\lambda} \mid \boldsymbol{f},\sigma^2,\mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{\lambda})\cdot p(\boldsymbol{\lambda} \mid \boldsymbol{f},\sigma^2). \tag{4.3.5}$$

Assume now that we observe the first M interfailure times with exponential distribution, then the likelihood function is

$$p(\mathcal{D} \mid \boldsymbol{\lambda}) = \prod_{i=1}^{M} \lambda_i e^{-t_i\lambda_i} \propto \exp\big(-\boldsymbol{\lambda}^T\boldsymbol{t}\big),$$

where $\mathcal{D} = \{t_i : i = 1,\ldots,M\}$ and $\boldsymbol{t} = (t_1,\ldots,t_M)^T$. From (4.3.5), the conditional posterior distribution for $\boldsymbol{\lambda}$ given the data and the parameters σ^2 and $\boldsymbol{f}$ is

$$p(\boldsymbol{\lambda} \mid \boldsymbol{f}, \sigma^2, \mathcal{D}) \propto \frac{1}{\left(2\pi\sigma^2\right)^{M/2}} \exp\left(-\boldsymbol{\lambda}^T \boldsymbol{t}\right) \exp\left(\frac{-1}{2\sigma^2}(\ln\boldsymbol{\lambda} - \boldsymbol{f})^T(\ln\boldsymbol{\lambda} - \boldsymbol{f})\right). \tag{4.3.6}$$

What remains is to sample the $\boldsymbol{f}$ and σ^2 from their conditional posterior distribution. Since we have assumed, $\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$, an inverse gamma as the prior distribution for σ^2, we have that

$$\begin{aligned} p(\sigma^2 \mid \boldsymbol{\lambda}, \boldsymbol{f}, \mathcal{D}) &\propto p(\boldsymbol{\lambda} \mid \boldsymbol{f}, \sigma^2) \cdot p(\sigma^2) \\ &\propto \frac{1}{\left(2\pi\sigma^2\right)^{M/2}} \exp\left(\frac{-1}{2\sigma^2}(\ln\boldsymbol{\lambda} - \boldsymbol{f})^T(\ln\boldsymbol{\lambda} - \boldsymbol{f})\right) \left(\sigma^2\right)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right) \\ &\propto \left(\sigma^2\right)^{-\left(\alpha+\frac{M}{2}\right)-1} \exp\left(\frac{-1}{\sigma^2}\left(\beta + \frac{1}{2}(\ln\boldsymbol{\lambda} - \boldsymbol{f})^T(\ln\boldsymbol{\lambda} - \boldsymbol{f})\right)\right), \end{aligned}$$

which we recognize as an inverse gamma,

$$\mathcal{IG}\left(\alpha + \frac{M}{2}, \beta + \frac{1}{2}(\ln\boldsymbol{\lambda} - \boldsymbol{f})^T(\ln\boldsymbol{\lambda} - \boldsymbol{f})\right), \tag{4.3.7}$$

and

$$\begin{aligned} p(\boldsymbol{f} \mid \boldsymbol{\lambda}, \sigma^2, \mathcal{D}) &\propto p(\boldsymbol{\lambda} \mid \boldsymbol{f}, \sigma^2) \cdot p(\boldsymbol{f}) \\ &\propto \exp\left(\frac{-1}{2\sigma^2}(\ln\boldsymbol{\lambda} - \boldsymbol{f})^T(\ln\boldsymbol{\lambda} - \boldsymbol{f})\right) \exp\left(\frac{-1}{2}\boldsymbol{f}^T C(\boldsymbol{\theta})^{-1}\boldsymbol{f}\right). \end{aligned}$$

Completing the square, we obtain the conditional posterior distribution of the corresponding function value $\boldsymbol{f}$,

$$p(\boldsymbol{f} \mid \boldsymbol{\lambda}, \sigma^2, \mathcal{D}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{f} - \boldsymbol{w})^T\left(\frac{1}{\sigma^2}\boldsymbol{I} + C(\boldsymbol{\theta})^{-1}\right)(\boldsymbol{f} - \boldsymbol{w})\right),$$

where $\boldsymbol{w} = \sigma^{-2}\left(\sigma^{-2}\boldsymbol{I} + C(\boldsymbol{\theta})^{-1}\right)^{-1} \ln\boldsymbol{\lambda}$, and we recognize the form of the posterior distribution as Gaussian with mean $\boldsymbol{w}$ and covariance matrix A^{-1},

$$p(\boldsymbol{f} \mid \boldsymbol{\lambda}, \sigma^2, \mathcal{D}) \propto \mathcal{N}\left(\sigma^{-2}A^{-1}\ln\boldsymbol{\lambda}, A^{-1}\right), \tag{4.3.8}$$

where $A = \sigma^{-2}\boldsymbol{I} + C(\boldsymbol{\theta})^{-1}$.

Next, we assume that $\mathcal{D} = \{n_i : i = 1,\ldots,M\}$ are the number of failures observed in time periods of length $L_1,\ldots,L_M$ which follows a Poisson distributed, then the likelihood is not a function of $\boldsymbol{\theta}$,

$$p(\mathcal{D} \mid \boldsymbol{\lambda}) = \prod_{i=1}^{M} \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_i}.$$

The prior of $\boldsymbol{f}$ is assumed to correspond to a GP with zero mean, and there exits a covariance function. An example of such a covariance function is given in (4.3.2). Then the distribution of $\boldsymbol{\lambda} \mid \boldsymbol{f}, \sigma^2$ is the same that in (4.3.3) and the prior distribution for σ^2 is an inverse-gamma, $\mathcal{IG}(\alpha, \beta)$. As before, we can evaluate the posterior distributions as follows. The joint posterior for $\boldsymbol{\lambda}, \boldsymbol{f}$ and σ^2 given the observed data is given in (4.3.4). We need to use numerical integration techniques or simulation methods to generate samples from these posterior distributions. Note that, from (4.3.5), we get

$$p(\boldsymbol{\lambda} \mid \boldsymbol{f}, \sigma^2, \mathcal{D}) \propto \frac{1}{\left(2\pi\sigma^2\right)^{M/2}} \exp\left(-\boldsymbol{\lambda}^T \mathbf{1} - \frac{1}{2\sigma^2}(\ln\boldsymbol{\lambda} - \boldsymbol{f})^T(\ln\boldsymbol{\lambda} - \boldsymbol{f})\right)\left(\prod_{i=1}^{M} \frac{\lambda_i^{n_i}}{n_i!}\right), \tag{4.3.9}$$

where $\mathbf{1}$ denotes a vector with all entries one, that is, $\mathbf{1} = (1, \ldots, 1)^T$. The conditional posterior distributions of σ^2 and $\boldsymbol{f}$ are given in (4.3.7) and (4.3.8), respectively.

4.3.3. Bayesian inference and model selection

Exact Bayesian inference is impossible for these models and thus, numerical methods must be applied. Various approaches have been developed for sampling in the context of GP models, see e.g. Rasmussen and Williams [80] and here we apply a MCMC algorithm suggested in Neal [70] in the context of non-parametric regression. The MCMC approach is the most popular and versatile method to evaluate the posterior distribution (see Subsection 4.2.1). Instead of an analytical expression for the posterior distribution, the method uses sampling data generated from this posterior distribution. The main idea behind the MCMC approach is to derive samples from the joint posterior distribution of the model parameters by alternatingly applying conditional marginal densities related to the joint posterior distribution. MCMC methods for NHPP based SRMs are discussed by Kuo and Yang [49], as we noted in Section 4.2.2.

Our scheme basically proceeds by specifying initial values for the parameters and then successively generating from their conditional posterior distributions using Gibbs or

Metropolis-Hastings steps as appropriate. Specifically, we consider the use of Gibbs sampling for conditionally conjugate distributions, i.e., for σ^2 given in (4.3.7) and for $\boldsymbol{f}$ given in (4.3.8), and the Metropolis-Hastings algorithm for distributions that are not conditionally conjugate, i.e., for $\boldsymbol{\lambda}$ given in (4.3.6) and (4.3.9) when the data are times between failures or number of failures, respectively. The acceptance rate in the Metropolis-Hastings step was about 35%.

Our proposed model class includes many simpler models such as the JM model (see Section 1.5) which are independent of covariate information and also simpler

regression functions. Furthermore, in many problems we may often have large numbers of metrics available and therefore, which model or which metrics to choose is an important problem. The standard approach to model selection in the classical context is to use selection criteria such as the Akaike or Bayesian information criterion. The most popular Bayesian selection criterion is the deviance information criterion or DIC developed in Speigelhalter et al. [93]. However, this criterion is highly dependent on the stability of the posterior (mean) parameter estimates and in the Gaussian process context, we have found that it is unstable. Therefore, we prefer to use a variant of the DIC, denoted DIC_3 in Celeux et al. [14]. This criterion is defined, for the Type II model with data $\boldsymbol{n} = (n_1, \ldots, n_M)$ and model $\mathcal{M}$ as

$$-4\,\mathbb{E}\left[\ln p(\boldsymbol{n}\boldsymbol{\theta}) \mid \boldsymbol{n}, \mathcal{M}\right] + 2\ln \widehat{p}(\boldsymbol{n} \mid \boldsymbol{n}, \mathcal{M}),$$

where

$$\widehat{p}(\boldsymbol{n} \mid \boldsymbol{n}, \mathcal{M}) = \prod_{i=1}^{M} \widehat{p}(n_i \mid \boldsymbol{n}, \mathcal{M}),$$

and

$$\widehat{p}(n_i \mid \boldsymbol{n}, \mathcal{M}) = \frac{1}{J}\sum_{j=1}^{J} p(n_i \mid \boldsymbol{n}, \lambda_{i,j}, \mathcal{M}) = \frac{1}{J}\sum_{j=1}^{J} \frac{\lambda_{i,j}^{n_i} e^{-\lambda_{i,j}}}{n_i!}.$$

This criterion is straightforward to calculate from the MCMC output and, in our experience, gives much more satisfactory results than the DIC. As with the AIC and BIC, lower values of this criterion imply better fitting models. For a full review, see Celeux et al. [14].

It is also interesting to study the predictive capacity of the GP based regression models. Rinsaka et al. [82] considered partitioning data set into two parts (50% and 75% and 90%) for predictive purposes, a training set for parameter estimation and a test set for evaluation of the predictive quality of their models. The *prediction square error* (PSE) is then calculated to evaluate the predictive ability of the models. For a training sample of size r, this is defined as

$$PSE = \frac{1}{M-r}\sum_{i=r+1}^{M}\Big(n_i - \mathbb{E}\left[N_i \mid n_1, \ldots, n_{i-1}\right]\Big)^2,$$

where the observed data are the numbers of software failures $N_1 = n_1, \ldots, N_M = n_M$. Clearly, lower values of the PSE indicate better predictive performance.

To study the predictive ability we compare our model with the model of Rinsaka et al. [82]. They developed a software reliability assessment tool, called PISRAT: Proportional Intensity-based Software Reliability Assessment Tool (see Shibata et al. [90]). We use this software program to obtain the prediction square error of the model by Rinsaka et al. [82] in order to compare the predictive performance of this model with

the predictive ability of our model. According to Rinsaka et al. [82], we denote by PIM($\cdot$,$\cdot$) the proportional intensity model which we outlined in Subsection 4.1.1, where the first component represents the software metrics used and the second is a distribution function, such as $\mathcal{W}$ = Weibull distribution, $\mathcal{G}$ = Gamma distribution and $\mathcal{E}$ = exponential distribution.

4.3.4. Applications to real data sets

In this subsection, we present the analysis of three real data sets. Two of them, referred to hereafter as DS1 and DS2 respectively, were taken from Rinsaka et al. [82] where alternative, non-Bayesian models were proposed. The third data set, referred to hereafter as DS3, was presented by Dalal and McIntosh [19]. All data sets consist of number of failures in given time periods and therefore can be analyzed using Type II models. DS1 and DS2 contain 54 and 38 failure counts observed during 17 and 14 weeks, respectively, along with the values of three testing effort data: execution time (CPU hr), failure identification work (person hr) and computer time-failure identification (CPU hr) which we shall treat as software metrics. DS3 contains approximately 400000 new or changed non-commentary source lines (NCNCSL), the staff time spent testing and the number of faults found. In order to undertake Bayesian inference for the models described before, prior distributions for the GP parameter σ^2 and hyperparameters $\boldsymbol{\theta} = (\rho_1^2, \ldots, \rho_k^2, \eta^2)$ must be defined. As is typical in such problems, we shall assume independent, proper but relatively uninformative inverse gamma, $\mathcal{IG}(\alpha, \beta)$, priors, where $\alpha = \beta = 0.001$. In all cases, our approach was implemented using the MCMC algorithm outlined in Subsection 4.3.3 with 10000 iterations to burn in and 50000 iterations in equilibrium.

RESULTS FOR DS1

Various models based on both linear regressions and on the use of Gaussian processes, with different software metrics as inputs, were considered for DS1. Table 4.1 shows the estimated values of the DIC_3 criterion from fitting various regression models for f (left hand side) and GP priors (right hand side) to these data. Note that many other Poisson regression models were also applied but performed worse than the optimal model presented here.

From Table 4.1, it can be seen that our models can give the smallest DIC_3, i.e., our models fit the data reasonably well. The optimal model according to the DIC_3 criterion is the GP model that uses all three software metrics, although various GP based models give similar results. The estimated mean number of failures and a 95% credible interval given the GP(EFC) model for the 17 weeks of DS1 are illustrated in

Table 4.1: DIC_3 criterion for different type II models for DS1

Model	DIC_3	Model	DIC_3
$\beta_0 + \beta_C x_C$	88.9387	GP(C)	59.5539
$\beta_0 + \beta_E x_E$	86.5568	GP(E)	59.3764
$\beta_F x_F$	67.9521	GP(F)	58.8286
$\beta_0 + \beta_E x_E + \beta_C x_C$	88.6426	GP(EC)	58.9651
$\beta_E x_E + \beta_F x_F$	69.2479	GP(EF)	58.7668
$\beta_F x_F + \beta_C x_C$	68.5449	GP(FC)	59.0388
$\beta_E x_E + \beta_F x_F + \beta_C x_C$	70.1835	GP(EFC)	58.5152

Figure 4.1. The model appears to estimated the observed data very well and in all cases, the observed value falls within the 95% predictive interval.

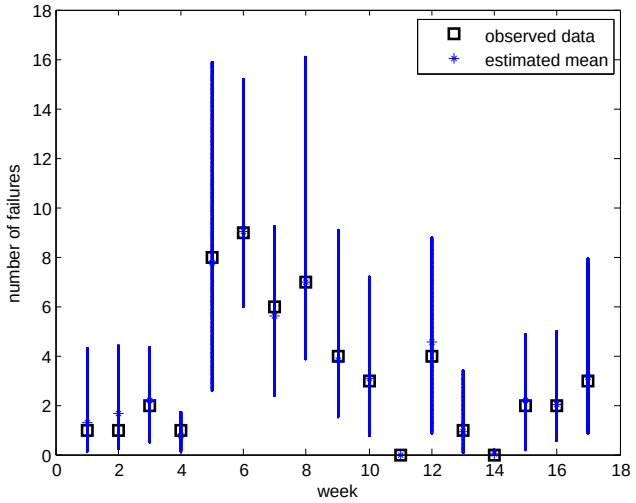


Figure 4.1: Estimated mean number of failures and 95% intervals for DS1.

In order to study the predictive capacity of the GP model, we perform one step prediction and following Rinsaka et al. [82], we shall consider training sets consisting of (approximately) the first 50% and 75% and 90% of the sample, that is 9, 13 and 15 data respectively so that the test sets consist of 8, 4 and 2 data, respectively. We then measure the predictive capacity of our models using the prediction squared error (PSE) in each case. Table 4.2 compares the predictive squared error performance of the three models with the smallest DIC_3s calculated earlier, that is GP(F), GP(EF) and GP(EFC) with the proportional intensity PIM$(\cdot,\cdot)$ model proposed by Rinsaka et al. [82], where the first component represents the metrics used (E = execution time, C = failure identification work and F = computer time-failure identification). In Table

4.2, one sees NC in some cases, it means that the algorithms of the PIM model do not converge, in spite of their computational efficiency (see Shibata et al. [90]).

Table 4.2: Prediction squared errors multiplied by 100 for DS1

Model	50%	75%	90%
PIM(F, $\mathcal{G}$)	143	1.0	2.0
PIM(F, $\mathcal{W}$)	7	2.0	4.0
GP(F)	4.73	2.53	1.60
PIM(EF, $\mathcal{W}$)	11.0	2.0	3.0
GP(EF)	8.89	1.6	3.65
PIM(EFC, $\mathcal{W}$)	NC	2.0	1.0
GP(EFC)	12.51	5.16	0.25

As seen from Table 4.2, the GP and PIM models show similar PSE performance for this data set. Note that many other PIM models were also applied but performed worse than the models presented in Table 4.2. Finally, Figure 4.2 shows the behavior of predicted number of faults from the 75% observation point with DS1 of the three GP based models with the smallest DIC_3s.

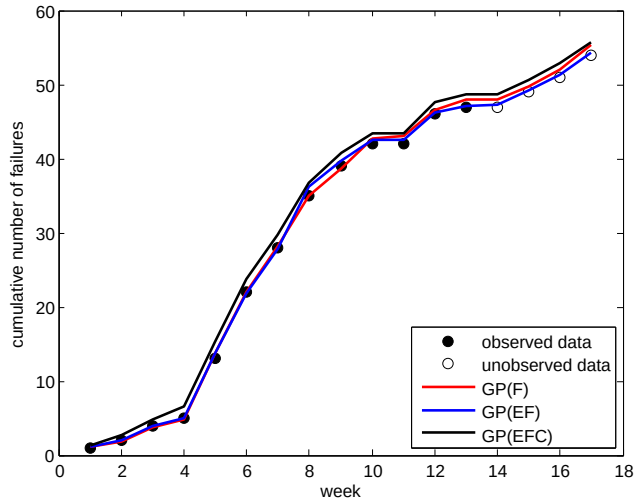


Figure 4.2: Predicted number of failures based on test set 2 (75%) for DS1.

The predicted number of faults appear reasonable.

RESULTS FOR DS2

Here we use the model with the smallest DIC_3 value, that is, GP(EFC), to compute the estimated mean number of failures and a 95% credible interval for the 14 weeks

of DS2. Figure 4.3 shows that this model estimates reasonably well the observed data, and also, the observed value in all weeks falls within the 95% predictive interval.

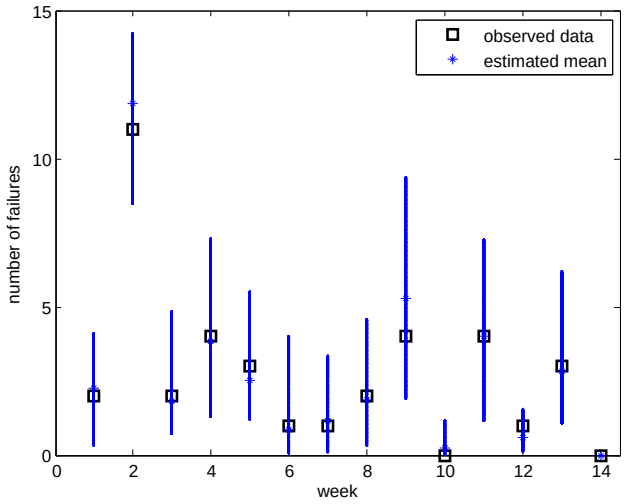


Figure 4.3: Estimated mean number of failures and 95% intervals for DS2.

Another three experiments have been performed using data DS2. In this case, the data were divided into training sets of size 7, 10 and 12 and test sets of size 7, 4 and 2 respectively. We measure the predictive capacity of the three models with the smallest DIC_3 in Table 4.1, and then, we compare these models with the PIM model. Table 4.3 presents the prediction error results multiplied by 100 for DS2. As before, NC means that the algorithms of the PIM model do not converge.

Table 4.3: Prediction squared errors multiplied by 100 for DS2

Model	50%	75%	90%
PIM(F, $\mathcal{E}$)	227	12	4
PIM(F, $\mathcal{G}$)	15	23	5
GP(F)	3.71	9.06	3.88
PIM(EF, $\mathcal{W}$)	475	32	3
PIM(EF, $\mathcal{G}$)	2702	15	2
GP(EF)	20.71	2.31	1.53
PIM(EFC, $\mathcal{W}$)	NC	36	4
PIM(EFC, $\mathcal{E}$)	NC	10	4
GP(EFC)	16.61	3.31	1.80

For DS2, it is clear that our model is better than PIM model. The largest difference can be seen in the first experiment from 50% observation point. However, the pre-

diction of both models is very similar in the third experiment where the observation point is 90%. Figure 4.4 graphs the cumulative number of failures for the real data DS2 and gives the GP regression prediction from the 75% observation point.

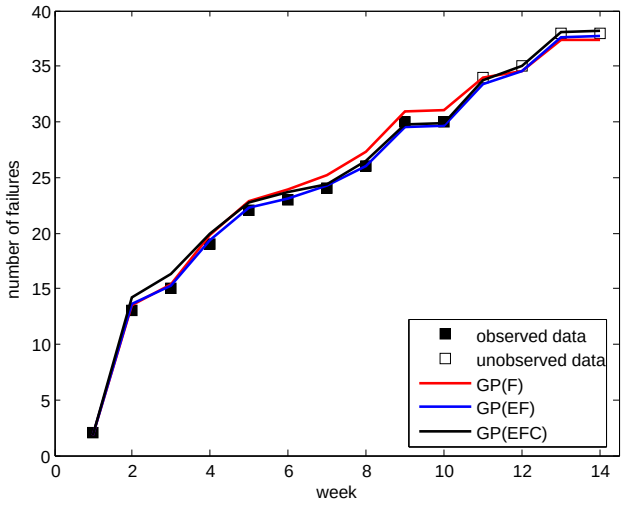


Figure 4.4: Predicted number of failures based on test set 2 (75%) for DS2.

RESULTS FOR DS3

We shall consider three training sets for DS3 consisting of 99 (50%), 149 (75%) and 178 (90%) data, and three test sets consist of 99, 49 and 20 data, respectively. We then compute the estimated values of the deviance information criterion of our model using the new or changed noncommentary source lines (NCNCSL) as covariate. In order to study whether software metrics provide information to the model, we compare the GP model with two classical NHPP-SR models defined in Subsection 1.5.2. In particular, we shall consider a Bayesian approach to the GO model and the DU model as follows,

$$\begin{aligned} N_i \mid a, b &\sim \mathcal{P}(\Lambda(t_i)) \\ a &\sim \mathcal{G}(\alpha_a, \beta_a) \\ b &\sim \mathcal{G}(\alpha_b, \beta_b), \end{aligned}$$

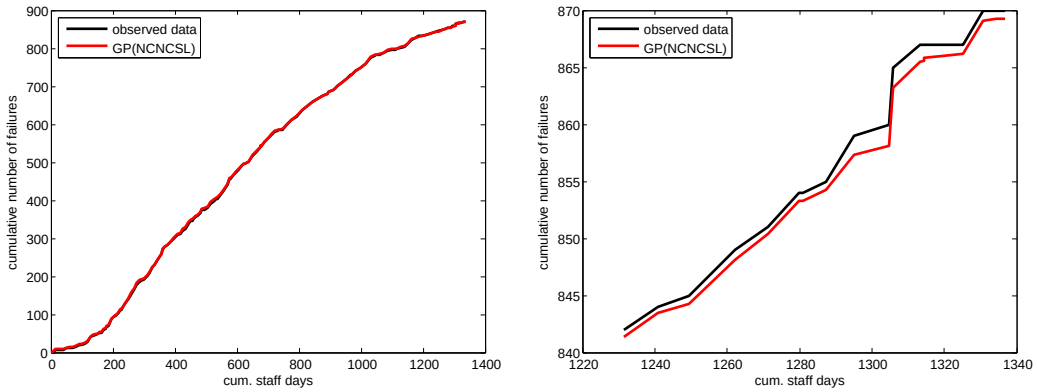
where $\Lambda(t)$ is defined in (1.5.13) for the GO model and in (1.5.15) for the DU model.

From Table 4.4, it can be seen that our model can give the smallest DIC_3 value, i.e., in the estimation of software failure data is appropriate to use software metrics information. We also perform one step prediction of the GP model using the software metric NCNCSL. Figures 4.5a, 4.6a and 4.7a plot the cumulative number of failures, observed and predicted, against the cumulative staff days for the three experiments,

Table 4.4: DIC_3 criterion for DS3

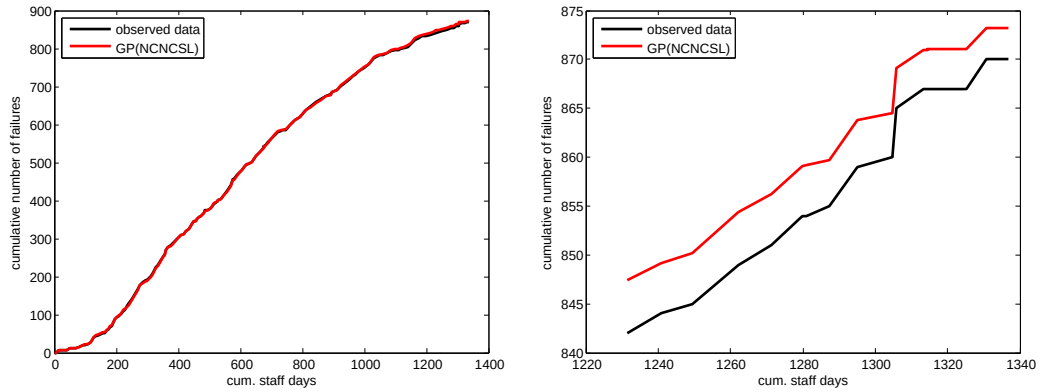
Model	50%	75%	90%
GP(NCNCSL)	299.80	517.90	625.73
GO-SRM	690.81	1.0786e+003	1.2929e+003
DU-SRM	694.12	1.0744e+003	1.3161e+003

50%, 75% and 90% observation point, respectively. Figures 4.5b, 4.6b and 4.7b plot the last 20 cumulative number of failures, observed and predicted, for the three experiments, respectively.



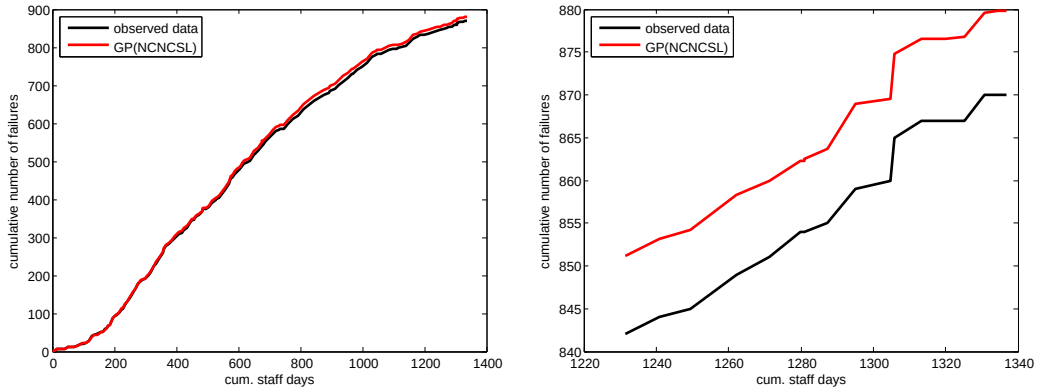
(a) full data set (b) the last 20 data

Figure 4.5: Predicted number of failures for DS3 and 50% observation point



(a) full data set (b) the last 20 data

Figure 4.6: Predicted number of failures for DS3 and 75% observation point



(a) full data set

(b) the last 20 data

Figure 4.7: Predicted number of failures for DS3 and 90% observation point

4.4. Conclusions and Extensions

In this chapter, we have shown how to combine software metrics data with failure counts data to improve the predictions of failure numbers of a software program using a Bayesian approach via Gaussian processes.

A number of extensions are possible. Firstly, in this paper we have considered the case were we have metrics and fault data and as we noted earlier, it is straightforward to extend the new type II model to the case of type I models.

Secondly, an important practical problem is what to do when not all metrics are available at all time periods. A reasonable approach to this would be to consider putting a prior distribution over the possible metric values when, given the observed data, the missing values could be easily sampled from the posterior distribution and incorporated within the MCMC algorithm outlined here.

Thirdly, many alternative approaches to modeling the function f could be considered. One possibility is to use spline functions, see e.g. Van der Linde [98] and another alternative would be to apply neural networks, see e.g. Su and Yuang [94] for a recent application in the software reliability context.

Work is currently underway on the above extensions.

CHAPTER 5

Conclusions and contributions of this thesis

This chapter describes the benefits arising from this thesis and the contributions that we have made to the fields of software reliability and stochastic orderings. The purpose of this thesis, as stated in the introductory chapter, was to study under what conditions a software system is reliable from a statistical point of view. To do this we have showed the relationship between the random variables that define software failure times, namely, times to software failure and interfailure times, and some statistic models, such as counting processes and models of ordered random variables. In the following we summarize the contributions of this thesis.

- We have studied stochastic properties of spacings from order statistics of heterogeneous exponential random variables, due to the relationship between spacings and the times elapsed between successive software failures.
  - In the one sample problem, we have shown that the conjecture of K&K [43] is true for $n = 4$ and we have established hazard rate ordering between the second and third normalized spacings. We also have obtained these results for simple spacings.
  - We have derived the conditions under which the spacings (both, simple and normalized) from two heterogeneous exponential random samples are ordered according to the likelihood ratio order. We have illustrated these results with an application to multiple-outlier models.
- Motivated by the relationship between sequential order statistics and the first n

epoch times of a nonhomogeneous pure birth process, we have studied stochastic properties and ageing notions of SOS.

- ➟ We have investigated the preservation of some ageing notions, such as IHR, IHRA, DHRA, from the underlying distribution function to the SOS.
- ➟ We have investigated conditions on the underlying distribution functions on which the SOS are based, in order to obtain stochastic comparisons between successive SOS.
- ➟ We have applied the results obtained for SOS to epoch times of NHPB processes.
- ➟ We have tried to developed a NHPB process approach to software reliability modelling under classical framework, but in this case the NHPB-SR model reduces to a NHPP-SR model.

➲ We have developed an alternative approach to software reliability models based on nonhomogeneous Poisson processes with covariate information in the form of software metrics such as code length, execution times will be generated each time the software is corrected.

- ➟ We have defined these new models which are useful to predict software failure for both Type I and Type II software reliability models.
- ➟ The models were constructed under Bayesian framework and the posterior inference was performed using Markov Chain Monte Carlo methods. The conditional approximation was implemented for Matlab which provides an efficient user interface and a wide variety of ready made toolboxes.
- ➟ Three real data case studies has been presented to illustrate the methodology developed.

Future research plans and extensions of the work presented in this thesis are included in each chapter, specifically in Sections 2.4, 3.7 and 4.4.

Bibliography

[1] B.C Arnold,, N. Balakrishnan, and H.N. Nagaraja. *Records*. Wiley Series in Probability and Statistics, 1998.

[2] I. Bagai and S.C. Kochar. On tail ordering and comparison of failure rates. *Communications in Statistics: Theory and Method*, 15:1377–1388, 1986.

[3] N. Balakrishnan and C.R. Rao. *Handbook os Statistics 16 - Order Statistics: Theory and Methods*. Elsevier, New York, 1998.

[4] N. Balakrishnan and C.R. Rao. *Handbook os Statistics 17 - Order Statistics: Applications*. Elsevier, New York, 1998.

[5] R.E. Barlow and F. Proschan. Inequalities for linear combinations of order statistics from restricted families. *Ann. Math. Statis.*, 38:1574–1592, 1966.

[6] J. Bartoszewicz. Dispersive functions and stochastic orders. *Applicationes Mathematicae(Warsaw)*, 24:429–444, 1997.

[7] L.A. Baxter. Reliability applications ofthe relevation transform. *Naval Research Logistics Quarterly*, 29:323–330, 1982.

[8] F. Belzunce, R. Lillo, and J.M. Ruiz. Stochastic comparisons of nonhomogeneous processes. *Probab. Eng. Inform. Sci.*, 15:199–224, 2001.

[9] J.R. Berrendero and J. Cárcamo. Tests for the second order stochastic dominance based on l-statistics. *Journal of Business and Economic Statistics*, 29:260–270, 2011.

[10] P.J. Boland, M. Shaked, and J.G. Shanthikumar. Stochastic ordering of order statistics. In N. Balakrishnan and C.R. Rao, editors, *Handbook os Statistics 16 - Order Statistics: Theory and Methods*, pages 89–103. Elsevier, New York, 1998.

[11] P.J. Boland and H. Singh. A birth-process approach to Moranda's geometric software reliability model. *IEEE Transactions on Reliability*, 52:168–174, 2003.

[12] M.C. Bryson, and M.M. Siddiqui. Some criteria for ageing. *J. Amer. Statist. Assoc*, 64:1472–1483, 1969.

[13] M. Burkschat, U. Kamps, and M. Kateri. Sequential order statistics with an order statistics prior. *Journal of Multivariate Analysis*, 101:1826–1836, 2010.

[14] G. Celeux, F. Forbes, C.P. Robert, and D.M. Titterington. Deviance information criteria for missing data models. *Bayesian Analysis*, 1:651–706, 2006.

[15] E. Cramer and U. Kamps. Sequential order statistics and *k*-out-of-*n* systems with sequentially adjusted failure rates. *Ann. Ins. Statist. Math.*, 48:535–549, 1996.

[16] E. Cramer and U. Kamps. Sequential k-out-of-n systems with Weibull components. *Econom. Quality Control*, 13:227–239, 1998.

[17] E. Cramer and U. Kamps. Estimation with sequential order statistics from exponential distributions. *Ann. Ins. Statist. Math.*, 53:307–324, 2001.

[18] E. Cramer and U. Kamps. Marginal distributions of sequential and generalized order statistics. *Metrika*, 58:293–310, 2003.

[19] S.R. Dalal and A.A. McIntosh. When to stop testing for large software systems with changing code. *IEEE Trans. Software Engineering*, 20:318–323, 1994.

[20] P. Deheuvels. The characterization of distributions by order statistics and record values - a unified approach. *J. Appl. Prob*, 21:326–334, 1984.

[21] J.T. Duane. Learning curve approach to reliability monitoring. *IEEE Trans. Aerospace*, 2:563–566, 1964.

[22] N.E.. Fenton and S.L. Pfleeger. *Software Metrics: A Rigorous & Practical Approach*. PWS Publishing, Boston, 1998.

[23] M.S. Finkelstein. On the reversed hazard rate. *Reliability Engineering & System Safety*, 78:71–75, 2002.

[24] J. Fox. *Nonparametric Simple Regression: Smoothing Scatterplots*. Thousand Oaks: Sage Publications, 2000.

[25] A.L. Goel and K. Okumoto. An analysis of recurrent software failures on a real-time control system. In *Proc. of the ACM Annual Technical Conference*, pages 496–500, 1978.

[26] A.L. Goel and K. Okumoto. Time-dependent error detection rate model for software reliability and other performance measures. *IEEE Trans. Reliability*, R, 1979.

[27] M. Grottke, l. Li, K. Vaidyanathan, and K.S. Trivedi. Analysis of software aging in a web server. *IEEE Transactions on Reliablity*, 55:411–420, 2006.

[28] R.C. Gupta and S.N.U.A. Kirmani. Closure and monotonicity properties of non-homogeneous poisson processes and record values. *Probability in Engineering and Information Sciences*, 2:475–484, 1988.

[29] M.H. Halstead. *Elements of software science*. Elsevier, New York, 1977.

[30] T. Hirata, H. Okamura, and T. Dohi. A Bayesian inference tool for NHPP-based software reliability assessment. In *Future Generation Information Technology, First International Conference, FGIT*, pages 225–236, 2009.

[31] T. Hu, Q. Lu, and S. Wen. Some new results on likelihood ratio orderings for spacings of heterogeneous exponential random variables. *Communications in Statistics: Theory and Methods*, 37:2506–2515, 2008.

[32] T. Hu, F. Wang, and Z. Zhu. Stochastic comparisons and dependence of spacings from two samples of exponential random variables. *Communications in Statistics*, 35:979–988, 2006.

[33] Z. Jelinski and P. Moranda. Software reliability research. In W. Freiberger, editor, *Statistical Computer Performance Evaluation*, pages 465–484. Academic Press, New York, 1972.

[34] U. Kamps. A concept of generalized order statistics. *J. Statist. Plann. Inference*, 48:1–23, 1995.

[35] U. Kamps. *A Concept of Generalized Order Statistics*. Teubner, Stuttgart, 1995.

[36] J. Keilson and U. Sumita. Uniform stochastic ordering and related inequalities. *Canadian Journal of Statistics*, 10:181–198, 1982.

[37] B. Khaledi and S.C. Kochar. Stochastic properties of spacings in a single-sutlier exponential model. *Probability in Engineering and Information Sciences*, 15:401–408, 2001.

[38] B. Khaledi and S.C. Kochar. Stochastic orderings among order statistics and sample spacings. In J.C. Misra, editor, *Uncertainty and Optimality: probability, statistics and operations research*, pages 167–203. World Scientific Publications, Singapore, 2002.

[39] S.C. Kochar. Some partial ordering results on record values. *Communications in Statistics:Theory and Methods*, 19:299–306, 1990.

[40] S.C. Kochar. Some results on interarrival times of nonhomogeneous poisson processes. *Probability in the Engineering and Informational Sciences*, 10:75–85, 1996.

[41] S.C. Kochar. Lorenz ordering of order statistics. *Statist. Probab. Lett.*, 76:1855–1860, 2006.

[42] S.C. Kochar and S.N.U.A. Kirmani. Some results on normalized spacings from restrictited families of distributions. *J. Statis. Plann.*, 46:79–57, 1995.

[43] S.C. Kochar and R. Korwar. Stochastic orders for spacings of heterogeneous exponential random variables. *J. Multivariate Anal.*, 57:69–83, 1996.

[44] S.C. Kochar and J. Rojo. Some new results on stochastic comparisons of spacings from heterogeneous exponential distributions. *J. Multivariate Anal.*, 57:69–83, 1996.

[45] S.C. Kochar and M. Xu. Stochastic comparisons of spacings from heteorgeneous samples. In M. Wells and A. Sengupta, editors, *Advances in Directional and Linear Statistics*, pages 113–129. Festschrift Volume for J.S. Rao, Springer, 2011.

[46] M. Krakowski. The relevation transform and a generalization of the gamma distribution function. *Rev. Française Automat. Inform. Rech. Opér. Sér. Verte*, 7:107–120, 1973.

[47] C. Kundu, A.K. Nanda, and T. Hu. A note on reversed hazard rate of order statistics and record values. *Journal of Statistical Planning and Inference*, 139:1257–1265, 2009.

[48] L. Kuo and T.Y. Yang. Bayesian computation of software reliability. *J. Comput. Graphical Stat.*, 4:65–82, 1995.

[49] L. Kuo and T.Y. Yang. Bayesian computation for nonhomogeneous poisson processes in software reliability. *J. Amer. Statist. Assoc*, 91:763–773, 1996.

[50] N. Langberg and N.D. Singpurwalla. unification of some software reliability models. *SIAM J. Sci. Stat. Comput.*, 6:781–790, 1985.

[51] K.S. Lau and B.L.S. Prakasa Rao. Characterization of the exponential distribution by the revelation transform. *J. Appl. Probab.*, 27:726–729, 1990.

[52] B. Lenz. Markovian simple counting processes and models ordered random variables. *Dissertation*, 2008.

[53] B. Littlewood and J.L. Verrall. A bayesian reliability growth model for computer software. *Appl. Stat.*, 22:332–346, 1973.

[54] M.R. Lyu. *Handbook of Software Reliability Engineering*. IEEE Computer Society Press, McGraw-Hill, 1996.

[55] H.L. MacGillivray, and K.P. Blanda. The relationships between skewness and kurtosis. *Australian Journal of Statistics*, 30:319–337, 1988.

[56] A.W. Marshall and I. Olkin. *Inequalities : Theory of Majorization and Its Applications*. Academic Press, New York, 1979.

[57] A.W. Marshall and I. Olkin. *Life distributions*. Springer, New York, 2007.

[58] R. Matias and P.J. Freitas. An experimental study on software aging and rejuvenation in web servers. In *Proc. 30th Annual International Computer Software and Applications*, pages 189–196, 2006.

[59] T.A. Mazzuchi and R. Soyer. A bayes empirical-bayes model for software reliability. *IEEE Trans. Rel.*, R, 1988.

[60] T.J. McCabe. A complexity measure. *IEEE Trans. Software Engineering*, 2:308–320, 1976.

[61] K. McDaid and S.P. Wilson. Determining an optimal time to test software assuming a time-dependent error detection rate model. *Technical Report 96/02, Department of statistics, Trinity College, Dublin,*, 1999.

[62] R.J. Meinhold and N.D. Singpurwalla. Bayesian analysis of a commonly used model for describing software failures. *The Statistician*, 32:168–173, 1983.

[63] D.R. Miller. Exponential order statistics models of software reliability growth. *IEEE Trans. Soft. Eng.*, 12:12–24, 1986.

[64] R. Misra, M. Manoharan, and H. Singh. Preservation of some aging properties by order statistics. *Probability in Engineering and Information Sciences*, 7:437–440, 1993.

[65] D.S. Mitrinovic. *Analytic Inequalities*. Springer-Verlag, Berlin, 1970.

[66] P.B. Moranda. Prediction of software reliability and its applications. In *Proc. of the Annual Reliability and Maintainability Symposium*, pages 327–332, 1975.

[67] J.D Musa and K. Okumoto. A logarithmic poisson execution time model for software reliability measurement. In *Proc. of the seventh International Conference on Software Engineering*, pages 230–237, 1984.

[68] H.N. Nagaraja. Some reliability properties of order statistics. *Communications in Statistics - Theory and Methods,*, 19:307–316, 1990.

[69] J. Navarro and M. Burkschat. Coherent systems based on sequential order statistics. *Naval Research Logistics*, 2011. To appear.

[70] R.M. Neal. Regression and classification using gaussian process priors (with discussion). In J. M. Bernardo and et al., editors, *Bayesian Statistics 6*, pages 475–501. Oxford University Press, 1998.

[71] I. Ntzoufras. *Bayesian modeling using WinBUGS*. Wiley Series in Computational Statistics, 2009.

[72] M. Ohba. Software reliability analysis models. *IBM Journal of Research Development,*, 21, 1984.

[73] E. Parzen. *Stochastic processes*. SIAM, Philadelphia, 1999.

[74] F. Pellerey, M. Shaked, and J. Zinn. Nonhomogeneous poisson processes and logconcavity. *Probab. Eng. Inform. Sci.*, 14:353–373, 2000.

[75] D. Pfeifer. Characterizations of exponential distributions by independent nonstationary record increments. *Journal of Applied Probability*, 19:127–135, 1982.

[76] D. Pfeifer. The structure of elementary pure birth processes. *Journal of Applied Probability*, 19:664–667, 1982.

[77] H. Pham. *System Software Reliability*. Springer, 2007.

[78] G. Pledger and F. Proschan. Comparisons of order satistics from heterogeneous populations, with applications in reliability. In J.S. Rustagi, editor, *Optimizing Methods in Statistics*, pages 89–113. Academic Press, New York, 1971.

[79] A.E. Raftery. Inference and rediction for a general order statistic model with unknown population size. *J. Amer. Statist. Assoc*, 82:1163–1168, 1987.

[80] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[81] B.K. Ray, Z. Liu, and N. Ravishankar. Dynamic reliability models for software using time-dependent covariates. *Technometrics*, 48:1–10, 2006.

[82] K. Rinsaka, K. Shibata, and T. Dohi. Proportional intensity-based software reliability modeling with time-dependent metrics. In *Proc. of the 30th IEEE Annual International Computer Software and Applications Conference (COMPSAC'06)*, 2006.

[83] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 2005.

[84] J. Rodrigues. Inference for the software reliability using asymmetric loss functions: A hierarchical Bayes approach. *Commun. Statist.-Theory Meth.*, 27:2165–2171, 1998.

[85] F. Ruggeri and R. Soyer. Advances in bayesian software reliability modelling. In T. Bedford, J. Quigley, L. Walls, B. Alkali, A. Daneshkhah, and G. Hardman, editors, *Advances in Mathematical Modeling for Reliability*. IOS Press, Netherlands.

[86] G.J. Schick and R.W. Wolverton. Assessment of software reliability. In *Proc. in Operations Research*, pages 395–422, 1978.

[87] M. Shaked. Dispersive ordering of distributions. *J. Appl. Probab.*, 19:310–320, 1982.

[88] M. Shaked and J.G. Shanthikumar. *Stochastic Orders*. Springer, New York, 2007.

[89] M. Shaked, F. Spizzichino, and F. Suter. Nonhomogeneous birth processes and l_∞-spherical densities, with applications in reliability theory. *Probab. Eng. Inform. Sci.*, 16:271–288, 2002.

[90] K. Shibata, K. Rinsaka, and T. Dohi. Pisrat: Proportional intensity-based software reliability assessment tool. In *Proc. of the 13th IEEE International Symposium on Pacific Rim Dependable Computing (PRDC'07)*, pages 43–52, 2007.

[91] N.D Singpurwalla, and S.P. Wilson. *Statistical Methods in Software Engineering*. Springer: New York, 1999.

[92] R. Soyer. Monitoring software reliability using non-Gaussian dynamic models. In *Proc. of the Engineering systems design and analysis Conference*, volume 1, pages 419–423, 1992.

[93] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64:583–639, 2002.

[94] Y.-S. Su and C.Y. Huang. Neural-network-based approaches for software reliability estimation using dynamic weighted combinational models. *Journal of Systems and Software*, 80:606–615, 2007.

[95] P.V. Sukhatme. Tests of significance for samples of the χ^2 population with two degrees of freedom. *Annals of Eugenics*, 8:52–56, 1937.

[96] N. Torrado, R.E. Lillo, and M.P. Wiper. Sequential order statistics: ageing and stochastic ordering. *Submitted*.

[97] N. Torrado, R.E. Lillo, and M.P. Wiper. On the conjecture of Kochar and Kowar. *J. Multivariate Anal.*, 101:1274–1283, 2010.

[98] A. Van der Linde. Splines from a bayesian point of view. *Tests*, 4:63–81, 1995.

[99] S. Wen, Q. Lu, and T. Hu. Likelihood ratio orderings of spacings heterogeneous exponential random variables. *J. Multivariate Anal.*, 98:743–756, 2007.

[100] M.P. Wiper. Software reliability: Bayesian analysis. In F. Ruggeri, R. Kennett, and W. Faltin, editors, *Encyclopedia of Statistics in Quality and Reliability*. Wiley, New York, 2007.

[101] M.P. Wiper, A.P Palacios, and J.M Marín . Bayesian software reliability prediction using software metrics information. *Quality Technology and Quantitative Management (in press)*, 2011.

[102] M.P. Wiper, D. Ríos Insua, and R. Hierons. Bayesian inference and optimal release times for two software failure models. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales*, 92:323–328, 1998.

[103] M.P. Wiper and M.T. Rodríguez Bernal. Bayesian inference for a software reliability model using metrics information. In *Proc. European Safety and Reliability: towards a safer world (ESREL'01)*, pages 1999–2006, 2001.

[104] S. Yamada and S. Osaki. Software reliability growth modeling: Models and applications. *IEEE Trans. Software Engineering*, 11:1431–1437, 1985.

[105] W. Zhuang and T. Hu. Multivariate stochastic comparisons of sequential order statistics. *Probab. Eng. Inform. Sci.*, 21:47–66, 2007.